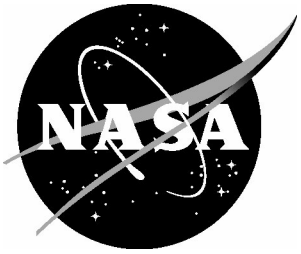
NASA/TP-2003-212442



Three Experiments Examining the Use of Electroencephalogram, Event-Related Potentials, and Heart-Rate Variability for Real-Time Human-Centered Adaptive Automation Design

Lawrence J. Prinzel III
Langley Research Center, Hampton, Virginia

Raja Parasuraman
Catholic University, Washington, D.C.

Frederick G. Freeman, Mark W. Scerbo, and Peter J. Mikulka
Old Dominion University, Norfolk, Virginia

Alan T. Pope
Langley Research Center, Hampton, Virginia

September 2003

The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at ***http://www.sti.nasa.gov***

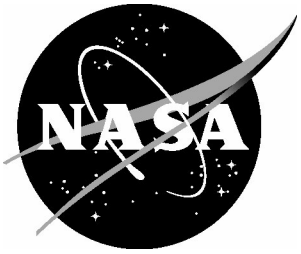
- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Phone the NASA STI Help Desk at (301) 621-0390

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

NASA/TP-2003-212442



Three Experiments Examining the Use of Electroencephalogram, Event-Related Potentials, and Heart-Rate Variability for Real-Time Human-Centered Adaptive Automation Design

Lawrence J. Prinzel III
Langley Research Center, Hampton, Virginia

Raja Parasuraman
Catholic University, Washington, D.C.

Frederick G. Freeman, Mark W. Scerbo, and Peter J. Mikulka
Old Dominion University, Norfolk, Virginia

Alan T. Pope
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

September 2003

The use of trademarks or names of manufacturers in the report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration or the U.S. Army.

Available from:

NASA Center for AeroSpace Information (CASI)
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161-2171
(703) 605-6000

Contents

Abstract 1
Introduction 1
  Impact of Automation Technology 1
    Advantages of Automation 1
    Disadvantages of Automation 2
Adaptive Automation 2
  Adaptive Mechanisms 3
Mental Workload 4
  Electroencephalogram 4
    Physiological Basis 4
    Description of the EEG 5
    Laboratory Studies 5
    Field Research 5
  Event-Related Potential 5
    Description 5
    Classification 6
    Physiological and Theoretical Basis 7
    Dual-Task ERPs 7
    Primary Task ERPs 8
    Simulation Research 9
    Conflicting Simulator Studies 10
    Real-Time Assessment of Mental Workload 11
  Heart-Rate Variability 12
    Cardiovascular Activity in Laboratory Tasks 12
    Cardiovascular Activity in Quasi-Operational Tasks 13
Research Purpose 14
  The Biocybernetic System 15
Experiment One 16
Method 17
  Participants 17
  Apparatus 18
  Experimental Tasks 18
  Experimental Design 18
  EEG Recording and Analysis 19
  Experimental Procedure 20
Results 21
  Task Allocations 21
  Electroencephalogram 21
  RMSE 21
  NASA-TLX 21
Discussion 22
  Task Allocations 22
  Electroencephalogram 22
  Tracking Performance 23
  Task Load 23

Experiment Two 24
Method 24
Participants 24
Apparatus 24
Experimental Design 25
Automation Cycle Sequencing 26
Dependent Variables 26
Statistical Tests and Criterion 26
Experimental Tasks 26
Tracking Task 26
Auditory Oddball Task 26
EEG Recording and Analysis 26
ERP Recording and Analyses 27
Experimental Procedure 28
Results 28
Task Allocations 28
Tracking Performance 28
Subjective Workload 29
Auditory Oddball Task Performance 29
Electroencephalogram 29
Event-Related Potentials 30
N100 Amplitude 30
N100 Latency 30
P200 Amplitude 30
P200 Latency 30
P300 Amplitude 32
P300 Latency 32
Discussion 32
Event-Related Potentials 33
Implications for Adaptive Automation 34
Mental Models 34
Applications to Adaptive Automation 34
Resource Allocation 35
Applications to Adaptive Automation 35
Experiment Three 36
Experiment 3a Method 36
Participants 36
Task 37
Heart Rate Recording 38
Procedure 38
Results and Discussion 38
MAT Performance 38
Heart Rate Measures 38
Experiment 3b Method 41
Participants 41
Task 41
Heart Rate-Based Adaptive Logic 41
Procedure 42

Results 42
Adaptive Transitions in Automation States 42
Adaptive System Evaluation: Heart Rate and MAT Performance 44
Discussion 45
Conclusions 46
Future Directions 48
References 49

Abstract

Adaptive automation represents an advanced form of human-centered automation design. The approach to automation provides for real-time and model-based assessments of human-automation interaction, determines whether the human has entered into a "hazardous state of awareness" and then modulates the task environment to keep the operator "in-the-loop", while maintaining an optimal state of task engagement and mental alertness. Because adaptive automation has not matured, numerous challenges remain, including what the criteria are, for determining when adaptive aiding and adaptive function allocation should take place. Human factors experts in the area have suggested a number of measures including the use of psychophysiology. This NASA Technical Paper reports on three experiments that examined the psychophysiological measures of event-related potentials, electroencephalogram, and heart-rate variability for real-time adaptive automation. The results of the experiments confirm the efficacy of these measures for use in both a developmental and operational role for adaptive automation design. The implications of these results and future directions for psychophysiology and human-centered automation design are discussed.

Introduction

Automation refers to "...systems or methods in which many of the processes of production are automatically performed or controlled by autonomous machines or electronic devices" (Parsons, 1985, p.7). Automation is a tool, or resource, that the human operator can use to perform some task that would be difficult or impossible without the help of machines (Billings, 1997). Therefore, automation can be thought of as a process of substituting some device or machine for some human activity; or it can be thought of as a state of technological development (Parsons, 1985). However, some people (e.g., Woods, 1996) have questioned whether automation should be viewed as a substitution of one agent for another. Nevertheless, the presence of automation has pervaded every aspect of modern life. We have built machines and systems that not only make work easier, more efficient and safer, but also have given us more leisure time. The introduction of automation has further enabled us to achieve these ends. With automation, machines can now perform many of the activities that we once had to do. Now, automatic doors open for us. Thermostats regulate the temperature in our homes, and automobile transmissions shift gears for us. We just have to turn the automation on and off. One day, however, there may not be a need for us to do even that.

Impact of Automation Technology

Advantages of Automation. Wiener (1980; 1989) noted a number of advantages to automating human-machine systems. These include increased capacity and productivity, reduction of small errors, reduction of manual workload and fatigue, relief from routine operations, more precise handling of routine operations, and economical use of machines. In an aviation context, for example, Wiener and Curry (1980) listed eight reasons for the increase in flight-deck automation: Increase in available technology, such as the Flight Management System (FMS), Ground Proximity Warning System (GPWS), Traffic Alert and Collision Avoidance System (TCAS); concern for safety; economy, maintenance, and reliability; decrease in workload for two-pilot transport aircraft certification; flight maneuvers and

navigation precision; display flexibility; economy of cockpit space; and special requirements for military missions.

Disadvantages of Automation. Automation also has a number of disadvantages. Automation increases the burdens and complexities for those responsible for operating, troubleshooting, and managing systems. Woods (1996) stated that automation is "...a wrapped package -- a package that consists of many different dimensions bundled together as a hardware/software system. When new automated systems are introduced into a field of practice, change is precipitated along multiple dimensions" (p.4). Some of these changes include: (a) adding to or changing the task, such as device setup and initialization, configuration control, and operating sequences; (b) changing cognitive demands, such as decreased situational awareness; (c) changing the role that people in the system have, often relegating people to supervisory controllers; (d) increasing coupling and integration among parts of a system often resulting in data overload and "transparency" (Billings, 1997); and (e) increasing complacency by those who use the technology. These changes can result in lower job satisfaction (automation seen as dehumanizing), lowered vigilance, fault-intolerant systems, silent failures, an increase in cognitive workload, automation-induced failures, over-reliance, increased boredom, decreased trust, manual skill erosion, false alarms, and a decrease in mode awareness (Wiener, 1989).

Adaptive Automation

These disadvantages of automation have resulted in increased interest in advanced automation concepts. One of these concepts is automation that is dynamic or adaptive in nature (Hancock & Chignell, 1987; Morrison, Gluckman, & Deaton, 1991; Rouse, 1977; 1988). In adaptive automation, control of tasks can be passed back and forth between the operator and automated systems in response to the changing task demands. Consequently, this allows for the restructuring of the task environment based upon (a) what is automated, (b) when it should be automated, and (c) how it should be automated (Rouse, 1988; Scerbo, 1996). Rouse (1988) described the criteria for adaptive aiding systems:

> The level of aiding, as well as the ways in which human and aid interact, should change as task demands vary. More specifically, the level of aiding should increase as task demands become such that human performance will unacceptably degrade without aiding. Further, the ways in which human and aid interact should become increasingly streamlined as task demands increase. Finally, it is quite likely that variations in level of aiding and modes of interaction will have to be initiated by the aid rather than by the human whose excess task demands have created a situation requiring aiding. The term *adaptive aiding* is used to denote aiding concepts that meet [these] requirements (p.432).

Adaptive aiding attempts to optimize the allocation of tasks by creating a mechanism for determining when tasks need to be automated (Morrison & Gluckman, 1994). In adaptive automation, the level or mode of automation can be modified in real-time. Further, unlike traditional forms of automation, both the system and the operator share control over changes in the state of automation (Scerbo, 1994; 1996). Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) have argued that adaptive automation represents the optimal coupling of the level of operator workload to the level of automation in the tasks. Thus, adaptive automation invokes automation only when task demands exceed the operator capabilities to perform the task(s) successfully. Otherwise, the operator retains manual control of the system functions. Although concerns have been raised about the dangers of adaptive automation (Billings & Woods, 1994; Wiener, 1989), it promises to regulate workload, bolster situational awareness, enhance vigilance, maintain manual skill levels, increase task involvement, and generally improve operator performance (Endsley, 1996; Parasuraman et al., 1992; Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1994, 1996; Singh, Molloy, & Parasuraman, 1993).

Adaptive Mechanisms

Perhaps, the most critical challenge facing system designers seeking to implement adaptive automation concerns how changes among modes or levels of automation will be accomplished (Parasuraman et al., 1992; Scerbo, 1996). The best approach involves the assessment of measures that index the operators' state of mental engagement (Parasuraman et al., 1992; Rouse, 1988). The question, however, is what should determine and "trigger" allocation of functions between the operator and the automation system. Numerous researchers have suggested that adaptive systems respond to variations in operator workload (Hancock & Chignell, 1987; 1988; Hancock, Chignell & Lowenthal, 1985; Humphrey & Kramer, 1994; Reising, 1985; Riley, 1985; Rouse, 1977), and that measures of workload be used to initiate changes in automation modes. Such measures include primary and secondary-task measures, subjective workload measures, and physiological measures. This, of course, presupposes that levels of operator workload can be specified so as to make changes in automation modes (Scerbo, 1996). Rouse (1977), for example, proposed a system for dynamic allocation of tasks based upon the operator's momentary workload level. Reising (1985) described a future cockpit in which pilot workload states are continuously monitored and functions are automatically reallocated back to the aircraft if workload levels get too high or too low. However, neither of these researchers provided specific parameters in which to make allocation changes (Parasuraman, 1990).

Morrison and Gluckman (1994), however, did suggest a number of workload candidates that may be used for initiating changes among levels of automation. They suggested that adaptive automation could be invoked through a combination of one or more real-time technological approaches. One of these proposed adaptive mechanisms is biopsychometrics. Under this method, physiological signals that reflect central nervous system activity, and perhaps changes in workload, would serve as a trigger for shifting among modes or levels of automation (Hancock, Chignell, & Lowenthal, 1985; Morrison & Gluckman, 1994; Scerbo, 1996).

Byrne and Parasuraman (1996) discussed the theoretical framework for developing adaptive automation around psychophysiological measures. The use of physiological measures in adaptive systems is based on the idea that there exists an optimal state of engagement (Gaillard, 1993; Hockey, Coles, & Gaillard, 1986). Capacity and resource theories (Kahneman, 1973; Wickens, 1984; 1992) are central to this idea. These theories posit that there exists a limited amount of resources to draw upon when performing tasks. These resources are not directly observable, but instead are hypothetical constructs. Kahneman (1973) conceptualized resources as being limited, and that the limitation is a function of the level of arousal. Changes in arousal and the concomitant changes in resource capacity are thought to be controlled by feedback from other ongoing activities. An increase in the activities (i.e., task load) causes a rise in arousal and a subsequent decrease in capacity. Kahneman's model was derived from research (Kahneman et al., 1967, 1968, 1969) on pupil diameter and task difficulty. Therefore, physiological measures have been posited to index the utilization of cognitive resources.

Several biopsychometrics have been shown to be sensitive to changes in operator workload suggesting them as potential candidates for adaptive automation. These include:

- Heart rate variability (Backs, Ryan, & Wilson, 1994; Itoh, Hayashi, Tsukui, & Saito, 1989; Lindholm & Cheatham, 1983; Lindqvist et al., 1983; Opmeer & Krol, 1973; Sayers, 1973; Sekiguchi et al., 1978)

- EEG (Natani & Gomer, 1981; O'Hanlon & Beatty, 1977; Sterman, Schummer, Dushenko, & Smith, 1987; Torsvall & Akerstedt, 1987)

- Eyeblinks (Goldstein, Walrath, Stern, & Strock, 1985; Sirevaag, Kramer, deJong, & Mecklinger, 1988)

- Pupil diameter (Beatty, 1982; 1986; 1988; Qiyuan, Richer, Wagoner, & Beatty, 1985; Richer & Beatty, 1985; 1987; Richer, Silverman, & Beatty, 1983)

- Electrodermal activity (Straube et al., 1987; Vossel & Rossmann, 1984; Wilson, 1987; Wilson & Graham, 1989) and

- Event-related potentials (Defayolle, Dinand, & Gentil, 1971; Gomer, 1981; Hancock, Chignell, & Lowenthal, 1985; Reising, 1985; Rouse, 1977; Sem-Jacobson, 1981).

There are several advantages to using biopsychometrics in adaptive systems. First, the measures can be obtained continuously with little intrusion (Eggemeier, 1988; Kramer, 1991; Wilson & Eggemeier, 1991). Second, because behavior is often at a low level when humans interact with automated systems, it is difficult to measure resource capacity with performance indices. Finally, these measures have been found to be diagnostic of multiple levels of arousal, attention, and workload. Therefore, it seems reasonable to determine the efficacy of using psychophysiological measures to allocate functions in an adaptive automated system. However, although many proposals concerning the use of psychophysiological measures in adaptive systems have been advanced, not much research has actually been reported (Byrne & Parasuraman, 1996). Nonetheless, many researchers have suggested that perhaps the three most promising psychophysiological indices for adaptive automation are the electroencephalogram (EEG), event-related potential (ERP), and heart-rate variability (HRV) physiological signal (Byrne & Parasuraman, 1996; Kramer, Trejo, & Humphrey, 1996; Morrison & Gluckman, 1994; Parasuraman, 1990; Scerbo, 1996).

Mental Workload

The use of psychophysiological measures in adaptive automation requires that such measures are capable of representing mental workload. Mental workload has been defined as the amount of processing capacity that is expended during task performance (Eggemeier, 1988). The basic concept refers to the difference between the processing resources available to the operator and the resource demands required by the task (Sanders & McCormick, 1993). Essentially, workload is invoked to describe the interaction between an operator performing the task and the task itself. In other words, the term "workload" delineates the difference between capacities of the human information processing system that are expected to satisfy performance expectations and that capacity available for actual performance (Gopher & Donchin, 1986).

Research has shown that the EEG, ERP, and HRV are useful as metrics of mental workload (Byrne & Parasuraman, 1996; Gale & Christie, 1987; Kramer, 1991; Parasuraman, 1990) and have unique properties that make them ideal for adaptive automation. A description of these three psychophysiological measures followed by a short review of these measures for mental workload assessment is presented next.

Electroencephalogram

Physiological Basis. The EEG derives from activity in neural tissue located in the cerebral cortex, but the precise origin of the EEG, what it represents, and the functions that it serves are not presently known. Current theory suggests that the EEG originates from post synaptic potentials rather than action potentials. Thus, the EEG is postulated to result primarily from the subthreshold post-synaptic potentials that

may summate and reflect stimulus intensity instead of firing in an all-or-none fashion (Gale & Edwards, 1983).

Description of the EEG. The EEG consists of a spectrum of frequencies between 0.5 Hz to 35 Hz (Surwillo, 1990). Delta waves are large amplitude, low frequency waveforms that typically range between 0.5 and 3.5 Hz in frequency, in the range of 20 to 200 *u*V (Andreassi, 1995). Theta waves are a relatively uncommon type of brain rhythm that occurs between 4 and 7 Hz at amplitude ranging from 20 to 100 *u*V. Alpha waves occur between 8 and 13 Hz at a magnitude of 20 to 60 *u*V. Finally, beta waves are an irregular waveform at a frequency of 14 to 30 Hz at amplitude of about 2 to 20 *u*V (Andreassi, 1995). An alert person performing a very demanding task tends to exhibit predominately low amplitude, high Hz waveforms (beta activity). An awake, but less alert person shows higher amplitude, slower frequency of activity (alpha activity). With drowsiness, theta waves predominate, and in the early cycles of deep slow wave sleep, delta waves are evident in the EEG waveform. The generalized effect of stress, activation or attention is a shift towards the faster frequencies, lower amplitudes with an abrupt blocking of alpha activity (Horst, 1987).

Laboratory Studies. Gale (1987) found that there exists an inverse relationship between alpha power and task difficulty. Other studies have also demonstrated the sensitivity of alpha waves to variations in workload associated with task performance. Natani and Gomer (1981) found decreased alpha and theta power when high workload conditions were introduced to pilots during pitch and roll disturbances in flight. Sterman, Schummer, Dushenko, and Smith (1987) conducted a series of aircraft and flight simulation experiments in which they also demonstrated decreased alpha power and tracking performance in flight with increasing task difficulty.

Numerous studies have also demonstrated that theta may be sensitive to increases in mental workload. Subjects have been trained to produce EEG theta patterns to regulate degrees of attention (Beatty, Greenberg, Diebler, & O'Hanlon, 1974; Beatty & O'Hanlon, 1979; O'Hanlon & Beatty, 1979; O'Hanlon, Royal, & Beatty, 1977). In particular, Beatty and O'Hanlon (1979) found that both college students and trained radar operators, who had been taught to suppress theta activity performed better than controls on a vigilance task. Though theta regulation has been shown to affect attention, the magnitude of the effect is often small (Alluisi, Coates, & Morgan, 1977). More recent research, however, has demonstrated its utility in assessing mental workload. Both Natani and Gomer (1981) and Sirevaag, Kramer, deJong, and Mecklinger (1988) found decreases in theta activity as task difficulty increased and during transitions from single to multiple tasks, respectively.

Field Research. More recent research has demonstrated the utility of EEG in assessing mental workload in the operational environment. Sterman et al. (1993) evaluated EEG data obtained from 15 Air Force pilots during air refueling and landing exercises performed in an advanced technology aircraft simulator. They found a progressive suppression of 8-12 Hz activity (alpha waves) at medial (Pz) and right parietal (P4) sites with increasing amounts of workload. Additionally, a significant decrease in the total EEG power (progressive engagement) was found at P4 during the aircraft turning condition for the air-refueling task (the most difficult flight maneuver). This confirmed other research that found alpha rhythm suppression as a function of increased mental workload (e.g., Ray & Cole, 1985).

Event-Related Potential

Description. The event-related potential, or ERP, is a transient series of voltage oscillations that occurs in response to the occurrence of a discrete event. This temporal relationship between the ERP and an event is what discriminates the ERP from the ongoing electroencephalogram (EEG) activity. The

ERP, like EEG, is a multivariate measure; however, unlike EEG, the ERP is broken down into a series of time rather than frequency domains (Kramer, 1991).

ERPs can be seen as a sequence of separate but often temporally overlapping components that are affected by a combination of the physical parameters of the stimuli and psychological constructs such as motivation, expectancy, resources, task relevance, memory, and attention (Kramer, 1987). Although the ERP has been found to be dependent upon both the psychological and physical characteristics of the eliciting stimuli, in some instances the ERP has been found to be independent of specific stimuli (Andreassi, 1995). For example, ERPs have been found to occur at the same time that the stimuli were *expected* to occur but were not actually presented (Sutton, Teuting, Zubin, & John, 1967).

Classification. The ERP can be classified as either being an evoked potential or an emitted potential. The "evoked potentials" (EPs) are ERPs that occur in response to physical stimulus presentation whereas "emitted potentials" occur in the absence of any invoking stimulus. Emitted potentials may be associated with a psychological process, such as recognition that a stimulus component is missing from a regular train of stimulus presentations or with some preparation for an upcoming perceptual or motor act (Picton, 1988).

ERP components can also be categorized along a continuum from endogenous to exogenous. The endogenous components are influenced by the processing demands imposed by the task, and are not very sensitive to changes in the physical parameters of stimuli, especially when these changes are not relevant to the task. In fact, endogenous components can be elicited by the absence of an eliciting stimulus if this "event" is relevant to the subject's task. Subject's strategies, expectancies, intentions, and decisions, in addition to task parameters and instructions, account for most of the endogenous components (Kramer, 1991).

The exogenous components, on the other hand, represent a response to the presentation of some discrete event. These components tend to occur somewhat earlier than endogenous components and they are usually associated with specific sensory systems, occur within 200 msec after the presentation of a stimulus, and are elicited by the physical characteristics of stimuli. For example, exogenous auditory potentials are influenced by the intensity, frequency, patterning, pitch, and location of the stimulus in the auditory field (Kramer, 1987; 1991).

The difference between the endogenous and exogenous components suggests the need for components to be clearly defined. ERP components are typically labeled with either a "N" or "P", for negative and positive polarity, respectively. Also, a number is assigned indicating the minimal latency measured from the onset of a discrete event. The attributes of the ERP that have served as definitional criteria have included: the arrangement of transient voltage changes across the scalp, polarity, latency range, sequence, and the sensitivity of these components to task instructions, parameters, and physical changes in the eliciting stimulus (Donchin, Ritter, & McCallum, 1978; Kramer, 1985; 1987; 1991).

The scalp arrangement concerns the amplitude and polarity of the components across various locations on the scalp. For example, research has demonstrated that the P300 component becomes increasingly smaller in amplitude from the parietal to the frontal sites, whereas the N100 is largest over the Fz, Cz, and Pz sites. The latency range is influenced by both experimental manipulations and whether it is an endogenous or exogenous component. For example, brainstem evoked potentials occur within 10 ms after the presentation of a stimulus. These ERPs are influenced by both organismic and stimulus variables; however, the latency range is only 2-5 ms. This is contrasted with the latency range of the P300 which depends on the processing requirements of the task and has been shown to span 300-900 ms (Kramer, 1991).

Physiological and Theoretical Basis. The ERP is composed of a sequence of "components" that are generated by groups of cells in different locations of the brain, which become active at different times after presentation of a stimulus. Although there is little consensus as to what the different components are thought to measure, the early components have been argued to represent the delivery of sensory input from various modalities through the afferent pathways. The later components originate in the primary projection systems, the different association areas, and the non-specific parietal and frontal regions (Vaughan & Arezzo, 1988).

To complicate matters further, the later the ERP components (e.g., P300), the more the components represent "memory-driven" rather than "data-driven" processes. For example, Hillyard and Picton (1979) have argued for a two-stage process for the ERP. The primary sensory system carries out a feature analysis and evaluates characteristics of the stimulus and, if it passes some criteria for selection, it then passes the sensory input to a second system. This second system evaluates the stimulus with comparison to memory models of expected or salient events (Gopher & Donchin, 1986).

The two-stage model of attentional processes involved in the etiology of the ERP has implication for the study of mental workload. Donchin and his colleagues (Donchin, 1981; Donchin, McCarthy, Kutas, & Ritter, 1983) argued that, because the P300 is elicited by improbable or unexpected events, the P300 represents a "context-updating" of the mental model of the environment. The mental model is continually assessed for deviations from expected sensory inputs and, when the events exceed some criterion, the mental model is updated. The frequency at which the mental model is updated is based on the surprise value and task relevance of the event. Donchin (1981) further developed a subroutine metaphor for the various activities of the ERP components. The P300 subroutine was posited to be invoked whenever there is a need to evaluate unusual, novel events in the environment (Gopher & Donchin, 1986; Kramer, 1987; Kramer; 1991).

The finding that the subroutine, characterized by the P300, is invoked only with task-relevant or surprising events has been important in the use of the ERP as a measure of mental workload. Consider a situation in which a participant must perform an oddball task while performing another task simultaneously. Now, imagine that the difficulty of the primary task is increased. Would the P300 subroutine still be invoked? If so, would the amplitude of the P300 reflect the increased workload demands and, therefore, serve as an index of the resources demanded by these two tasks? Such questions as these served as the impetus for researchers to begin to investigate the use of the P300 in the assessment of workload (Kramer, 1987; Gopher & Donchin, 1986; Parasuraman, 1990).

Dual-Task ERPs. The earlier ERP studies of mental workload were driven by research findings connecting changes in ERP components to state variables, such as fatigue and arousal. Haider, Spong, and Lindsley (1964) first reported that shifts in the N100 visual and auditory ERP during discrimination tasks reflected both states, such as fatigue, arousal, and vigilance, as well as discrimination task performance. Thereafter, ERPs were linked to the secondary-task method, a method that was emerging as a technique for assessing primary task workload demands. The earlier dual-task ERP studies of mental workload concentrated on stimulus-evoked, exogenous, rather than task-evoked, endogenous ERP components. For example, Defayolle, Dinand, and Gentil (1971) reported that the P100 component of the ERP to flashes of red light was reduced when subjects performed a reasoning task as opposed to a control condition in which no task was performed. Furthermore, as the difficulty of the reasoning task was increased, the amplitude of the P100 showed further reductions. Spyker, Stackhouse, Khalafall, and McLane (1971) demonstrated that the P250 component of the ERP was also affected by the difficulty of the task. They reported that the amplitude of the P250 component of the ERP to visual probe stimuli was reduced as the dynamic complexity of a tracking task was increased (Parasuraman, 1990).

In a recent review of the research, Parasuraman (1990) concluded that these early studies were plagued by lack of experimental control over the processing of the probe stimulus. The experimental tasks were either not integrated with the presentation of the probe or, as in the case of Defayolle, Dinand, and Gentil (1971), time domains of ERPs were not averaged separately for various response categories and different stimuli. More recent research, however, requires subjects to process the discrete event to some degree. A separate task is associated with the ERP stimuli making this method a more exact analog of the dual-task procedure (Parasuraman, 1990).

Many of these more recent studies have focused on the P300 component. These studies were based upon the notion that P300 amplitude in a task should be proportional to the attentional resources invested in the task (Johnson, 1986; Parasuraman, 1990). Put another way, if subjects are given one task to perform while performing another task concurrently, the demands imposed by the secondary task would impact the "memory-driven" processes and, therefore, can be assessed by evaluating how the amplitude of the P300 changes in the primary task (Parasuraman, 1990).

Wickens, Isreal, and Donchin (1977) reported one of the first studies to investigate the endogenous P300 component. In this study, the P300 amplitude to counted tones decreased when a visual tracking task was also performed. This finding is not much different than the earlier ERP studies, except that the effect was for a task-evoked, endogenous rather than a stimulus-evoked, exogenous ERP component. However, P300 amplitude was not found to be sensitive to increases in the difficulty of the tracking task, either when the number of tracked dimensions was increased from one to two (Wickens et al., 1977) or when the bandwidth of the tracking task was increased (Isreal, Chesney, Wickens, & Donchin, 1980). The fact that the P300 did not vary much as a function of primary task difficulty was attributed to the idea that primary and secondary tasks draw on different "resource pools." This view contends that the tracking task difficulty taps response-related resources; however, the P300 counting task taps perceptual resources.

In another study, Isreal, Wickens, Chesney, and Donchin (1980) coupled a counting task with a visual monitoring task. Subjects were asked to monitor the visual task for changes in the intensity or direction of squares and triangles that moved over a visual display. In this study, perceptual factors were manipulated by requiring subjects to monitor either four or eight display elements. The results showed that the P300 amplitude to the stimuli in the visual task was smaller in the dual-task conditions. Moreover, P300 was decreased further in the high-load, eight display element condition; however, this effect was found only for the direction-change primary task. Similar studies (e.g., Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Ragot, 1984) have also found that the P300 is influenced by perceptual factors. Taken together, these studies support the view that P300 amplitude can be used as a measure of workload of a perceptual and cognitive, but not response-related nature. Further, P300 latency has been found to change with stimulus parameters, such as masking, that are known to affect encoding and central processing, but not for stimulus-response processing, such as stimulus-response compatibility (McCarthy & Donchin, 1981; Parasuraman, 1990). These results have been discussed in terms of the multiple-resource view of workload that holds that several separate resource pools exist corresponding to different modalities, perceptual versus response processes, and so on (Wickens, 1984). The fact that the P300 amplitude was not sensitive to tracking difficulty suggests that this factor depletes resources that are not used by the P300 process (Hoffman, 1990; Parasuraman, 1990).

Primary Task ERPs. The afore-mentioned studies utilized a dual-task methodology to assess ERP as a metric to resources of a perceptual/cognitive nature and were taken as supporting the multiple-resource view of workload. The results demonstrated that, if the primary task difficulty is manipulated and yields secondary task performance decrements, in addition to secondary task P300 amplitude decrements, then

the results can be taken as reflecting competition for perceptual/central processing resources over and above those placed upon the response/output system. However, according to Sirevaag, Kramer, Coles, and Donchin (1989), the P300 associated with the primary task has been overlooked. They contended that, if P300 amplitude does indeed evince resource competition shown to occur during dual-task performance, logically then the P300s elicited by the primary task should result in an increase in amplitude as the workload of the primary task is increased. Further, in dual-task studies where ERPs can be recorded in response to both discrete primary and secondary task events, one should find a reciprocal relationship between primary and secondary task P300 amplitudes (Sirevaag et al., 1989).

The amplitude reciprocity hypothesis was tested in a study by Wickens, Kramer, Vanasse, and Donchin (1983) in which subjects were asked to track a target with a cursor. The ERPs elicited by the discrete changes of the primary task were recorded in one experimental run. ERPs for tones counted during the secondary task were also recorded in a separate trial. In this study, task demands were manipulating by changing the number of integrations between the joystick output and the movements of the cursor on the screen. They found that the P300 associated with the step changes increased in amplitude with increasing primary task difficulty; whereas secondary task P300 amplitudes decreased.

Recent studies have also found that P300s elicited to events from the primary task increase in amplitude with increases in primary task difficulty (Sirevaag et al., 1989; Strayer & Kramer, 1990; Ullsperger, Metz, & Gille, 1988). For example, Sirevaag et al. (1989) employed a method where both primary and secondary ERPs could be concurrently recorded within the same experimental condition. Measures of P300 amplitude and performance were obtained from 40 subjects within the context of a pursuit step-tracking task performed alone and with a concurrent secondary auditory discrimination task. The pursuit tracking task difficulty was manipulated by varying the velocity and acceleration control dynamics as well as the number of dimensions, either one or two, to be tracked. ERPs were recorded for both the tracking task setup changes and for the secondary task tones. The results showed that, as the primary task difficulty was increased as reflected in increased root mean squared error (RMSE) scores, there was decreased secondary task P300 amplitudes and increased primary task P300 amplitudes. Moreover, the increases in primary task P300 amplitudes were concomitant with the amplitude decrements obtained for the secondary task. These findings were taken as supporting the amplitude reciprocity hypothesis between primary and secondary task P300 amplitudes as a function of primary task difficulty.

Simulation Research. The previously mentioned research has provided important evidence about the relationship between the P300 and mental workload. However, these studies have not addressed whether such findings can generalize to real-world environments. This is especially important if such studies are to be applied to adaptively automated systems. Fortunately, much research has been conducted that has addressed this issue. Studies have employed a number of primary tasks, including pursuit and compensatory tracking, flight control and navigation, and memory/visual search, as well as both visual and auditory secondary tasks (Hoffman et al., 1985; Humphrey & Kramer, 1994; Kramer & Strayer, 1988; Kramer, Sirevaag, & Braune, 1987; Kramer, Wickens, & Donchin, 1983; 1985; Lindholm, Cheatham, Koriath, Longridge, 1984; Natani & Gomer, 1981; Sirevaag et al., 1993; Strayer & Kramer, 1990; Theissen, Lay, & Stern, 1986). For example, Lindhom et al. (1985) elicited ERPs to auditory stimuli during simulated landings and attack scenarios. They reported a larger P300 amplitude decrease as the workload in the primary task was increased. A related study used an oddball, or rare event, secondary-task to elicit ERPs as subjects performed a flight task simulation (Natani & Gomer, 1981). This study found significant P300 amplitude decrements as well as longer P300 latencies under the high workload conditions. However, similar results were not found for a second replication of the task (Wilson & Eggemeier, 1991).

Theissen, Lay, and Stern (1986) employed a visual oddball task to elicit ERPs while electronic warfare officers performed various tasks in a fighter aircraft simulator. Task difficulty levels were manipulated by changing task parameters, such as target characteristics (e.g., number and type) and threats to aircraft. The results demonstrated smaller P300 amplitudes in the single-task control condition than in the simulated flight conditions. Kramer, Sirevaag, and Braune (1987) evaluated workload during a flight simulation experiment that used an auditory, rather than visual, oddball task that required subjects to discriminate infrequent from frequent tones. They found that the P300 component of the ERP consistently indexed changes in flight difficulty level with a finding of decreased P300 amplitude with increased primary-task difficulty. Further, P300 amplitude demonstrated a negative correlation with deviations from flight headings. Such a finding suggests that primary task data can be coupled with ERP data to make allocation decisions in an adaptively automated environment.

Sirevaag et al. (1993) elicited ERPs to irrelevant probes as helicopter pilots flew a series of reconnaissance missions in a motion-based, high fidelity helicopter simulator. They reported smaller P300s amplitudes to probes as the communication load imposed on the pilots was increased. Biferno (1985) also looked at communication load and ERPs. He recorded ERPs from radio call signs as subjects performed flight simulator missions. P300 amplitude was found to be smaller as the workload increased. Furthermore, both fatigue and subjective workload estimates of workload were reported to discriminate between various levels of workload. These results suggest that ERPs are associated with other measures of taskload thereby attesting to their utility for workload estimation and adaptive automation.

Most of the research conducted with ERPs and mental workload has been focused on flight simulation. In one of the few applications of ERPs outside of aviation, Wesensten et al. (1993) recorded auditory ERPs from 10 male participants at 0900, 1600, and 1830 hours. P300s were collected while participants were at sea level and another one was collected following a rapid ascent to a simulated 4,300-meter altitude. The results of the study were a decrease in P300 amplitude, while P300 latency and reaction time increased, following the ascent. Another study (Janssen & Gaillard, 1985) used an auditory Sternberg memory task to elicit ERPs from automobile drivers as they drove on three different types of roadway: rural, city, and highway. Highway driving was found to elicit the smallest P300 amplitudes, and this was interpreted as being the driving segment with the highest workload (Wilson & Eggemeier, 1991).

Conflicting Simulator Studies. A number of field studies have demonstrated that the ERP reliably varies with workload. However, a few studies exist that have not shown such clear-cut evidence (e.g., Fowler, 1994; Jannsen & Gaillard, 1985; Natani & Gomer, 1981). For example, Fowler (1994) elicited ERPs using auditory and visual oddball tasks as subjects flew a final approach and landing manuever under workloads varied by manipulating turbulence and hypoxia. The oddball tasks required subjects to detect infrequent tones or flashes of an artificial horizon. Although RMSE flying performance was found to be systematically degraded by the two-workload conditions, the P300 amplitude was not strongly related to performance. However, P300 amplitude was inversely related to high taskload when the visual condition was analyzed separately. The authors accounted for this result by invoking the amplitude reciprocity hypothesis. As stated previously, this hypothesis suggests that, as the primary task difficulty is increased and the P300 amplitude elicited by the secondary task decreases, P300 amplitude for task-relevant events embedded in the primary task increases. Therefore, the flashing horizontal horizon was processed as part of the primary task causing the P300 amplitude to increase as a function of task difficulty. However, this cannot account for the results reported for the auditory condition as no systematic pattern emerged in contrast to a similar study done by Kramer, Sirevaag, and Braune (1987).

Fowler (1994) also reported that P300 latency was found to covary with flight performance, increasing as a function of workload in both modalities. O'Donnell and Eggemeier (1986) suggested that the P300

amplitude indexes workload because it is sensitive to subject expectancy that is disrupted by workload. This would explain the disassociation between latency and amplitude because the mechanisms controlling expectancy would be different than those indexing the speed of perceptual/cognitive processing. According to this view, the instrument flight rules (IFR) flying task used by Kramer, Sirevaag, and Braune (1987) primarily interrupted subject expectancy whereas the visual flight rules (VFR) task used by Fowler (1994) primarily slowed stimulus evaluation. The authors noted that this possibility suggests that both P300 amplitude and latency can be used as indices of mental workload, depending on the nature of the task (Fowler, 1994)

In a second study, Janssen and Gaillard (1985) were unable to replicate the finding of a smaller P300 amplitude to probes during expressway driving despite the fact that heart-rate variability was found to be significantly decreased in the more demanding expressway segment in both studies. Also, Natani and Gomer (1981) were unable to replicate the findings of their first study. Similar to Fowler (1994), however, Janssen and Gaillard reported that P300 latency was sensitive to increases in taskload.

Real-Time Assessment of Mental Workload. Although the simulator studies cited above, have yielded useful information, they have not addressed whether ERPs could measure dynamic changes in mental workload. For example, in simulator studies, 50-100 single trial ERPs may be collected and then averaged to determine whether ERP components discriminate workload or performance levels. In an adaptively automated environment, collection of this quantity of ERP data may not be practical. A number of earlier studies, however, have suggested that ERPs can be used for on-line evaluations of moment-to-moment fluctuations in operator workload (Defayolle et al., 1971; Gomer, 1981; Sem-Jacobsen, 1981). Although research on real-time assessment of mental workload is still in its infancy, this line of research has been expanded in several recent studies that have suggested that on-line assessment may soon be feasible. For example, Farwell and Donchin (1988) asked subjects to attend to one item in a 6 x 6 matrix of items. The columns and rows flashed randomly and ERPs elicited from the flashes were used to discriminate between the attended and unattended items. A 95 percent accuracy level was found using just 26 seconds of ERP data. Kramer, Humphrey, Sirevaag, and Mecklinger (1989) also found that on-line assessment of mental workload can be performed with a small amount of ERP data (Kramer, 1991).

Humphrey and Kramer (1994) also reported a study that examined whether ERPs could measure dynamic changes in mental workload. They examined how much ERP data is necessary to discriminate between levels of mental workload in complex, real-world tasks. In order to address this question, they employed a bootstrapping approach to investigate the accuracy of discriminating between workload levels using different amounts (e.g., 1 to 75 sec) of ERP data. Participants were asked to perform two tasks, monitoring and mental arithmetic, both separately and together. Following an analysis of the performance, subjective workload ratings, and average ERP data in the single- and dual-task conditions, two different conditions from each of the tasks were selected for further analysis. The results of the study indicated that 90% correct discrimination could be achieved with from 1 to 11 seconds of ERP data. These results were discussed in terms of real-time assessment of mental workload using ERP data. Kramer, Trejo, and Humphrey (1996) discussed these results as evidence that event-related potentials can be useful in the design of adaptive systems.

To conclude, the research on event-related potentials has consistently shown that the ERP can reliably and accurately measure the mental workload demands being imposed on the human operator. The ERP research has additionally demonstrated the advantage of the measure to characterize the quality of operator information processing, which would be of significant value in the monitoring of cognitive states in supervisory control environments. A disadvantage, however, of the ERP is the intrusiveness and

difficulty in implementing the method and the considerable expertise needed to interpret the results. Another psychophysiological measure that does not present such difficulties is the heart-rate variability (HRV), which is described next. However, although the HRV is a useful measure of cognitive workload, it does not have the same capability as the ERP in terms of its diagnosticity of information processing. Nevertheless, because of its ease of use and reliability, the HRV holds significant promise as a workload measure that could be easily implemented into an adaptively automated system.

Heart-Rate Variability

Cardiovascular activity is the most commonly used index of cognitive workload. It is a relatively unobtrusive physiological measure and it appears to be readily accepted by subjects in an operational environment. In a recent review of applied physiological measurement techniques, Fahrenberg and Wientjes (2000) ranked cardiovascular measurement as the most suitable for field studies due to its reliability, unobtrusiveness and ease of recording. Of the studies in this review, 21 used one or more indexes derived from heart activity, and many studies combined this with other physiological indexes. The earlier literature reports a consistent pattern of cardiovascular activity from laboratory and field studies; heart rate increases and heart rate variability decreases as a function of increases in cognitive workload (Wilson, 1992).

One trend in the use of cardiovascular function as a measure of workload, specifically mental workload, is the assertion that heart rate is not a sensitive or an especially diagnostic measure. There are two reasons for this. First, it is affected by physical exertion and second, it does not provide information about the underlying functioning of the sympathetic and parasympathetic nervous systems. Several authors feel that it is only through an understanding of the relative contributions of the autonomic nervous system on cardiovascular functioning that good diagnosticity of mental workload can be achieved (Backs, 1995; Berntson, Cacioppo, & Quigley, 1993; Jorna, 1992; Mulder, Mulder, Meijman, Veldman, & van Roon, 2000).

Spectral analysis of variations in heart rhythm is proposed to provide an index of the relative contributions of the underlying components: parasympathetic inhibition and sympathetic activation. Spectral analysis of heart rhythm is typically segmented into three distinct bandwidths: 1) low frequency (0.02-0.06Hz), which is associated with temperature regulation; 2) mid-frequency (0.07-0.14Hz), which is affected by blood pressure regulation and cognitive effort; 3) hi-frequency (0.15-0.50Hz) which is associated with the effects of respiration on heart rate, the respiratory sinus arrhythmia (RSA). The mid-frequency bandwidth is associated with the combined activity of the parasympathetic and sympathetic systems, while the RSA is influenced by parasympathetic activity. Mulder, et al. (2000) suggest that suppression of the mid-frequency bandwidth is "very diagnostic" of the operation of attention-demanding cognitive control mechanisms (i.e., mental workload). Another measure has been developed to reflect the impact of sympathetic activation on heart rhythm, residual heart rate (RHR). Residual heart rate is the heart rate that remains after removing the part linearly related to respiratory activity, RSA.

Cardiovascular activity in laboratory tasks. Boutcher, Nugent, McClaren and Weltman (1998) challenged aerobically fit men and two control groups with the Stroop task and an arithmetic task (subtraction of a series of spoken numbers). The premise for this study was that fit males have a greater vagal tone, increased parasympathetic activity, which may affect reactivity to mental challenge. Of relevance to the present review was the effect of the two cognitive tasks on cardiovascular function as measured by HRV in mid and high bands. The relevant comparison was between baseline and the given task. For the arithmetic task there were no significant changes for either HRV band, although there was a trend for a reduction in variability during the task. However, the same comparison of the Stroop task revealed a

significant reduction of HRV in both bands. Sammer (1998) compared a physical task (moving a lever when a cue appears), a cognitive task (counting target letters appearing in a serial array) and a combination of both task (dual task). Inter-Beat Interval (IBI), and HRV in the low (0.01-0.05Hz), mid (0.06-0.16Hz), and high (0.2-0.4Hz) bands were computed. A comparison among the three tasks (no baseline comparison was included) found significant effects for all four measures. Heart period was largest (slowest HR) for the cognitive task, intermediate for the physical task, and smallest for the dual task (faster HR). Over the spectral bands, HRV was less for the dual task and greater for the physical and cognitive tasks, which were not different. Simply, heart period differentiated among the tasks better than the HRV measures. Fournier, et al., (1999) used the Multiple Attribute Task Battery and created four discrete tasks: a single task and three multiple tasks of increasing difficulty. HR, and HRV in the mid and high bands were the dependent variables. In an initial comparison of the single task condition to the multiple tasks, all three measures were different: HR was higher and HRV in both bands was reduced in the multi-task conditions. A subsequent comparison among the three multiple tasks found that HR differentiated between the highest difficulty task (higher HR) and the other two multiple tasks, whereas only the mid-band HRV was different between the high and low difficulty multiple tasks.

The above studies suggest that the simple measure of HR was more sensitive and diagnostic that the HRV measure. Also, there was little evidence that the HRV mid band was more sensitive to mental challenges than the other spectral bands. Backs and his colleagues (Backs, 1995; 1997; Backs, Lenneman, & Sicard, 1999; Backs, Ryan, & Wilson, 1994) have proposed a complex decomposition of cardiovascular activity into autonomic dimensions (parasympathetic and sympathetic activity) in order to generate a more sensitive and diagnostic measure of workload. They conducted a series of studies using a single-axis, compensatory tracking task that varied physical demand by either: 1) requiring different amounts of force to move the joystick force, or 2) varying the disturbance value of the cursor movement, and varied cognitive/perceptual load by manipulating order-of-control (velocity, acceleration, mixed). Also, secondary tasks were added to increase discrete workloads (e.g., target recognition varying set size, mathematical tasks, oddball counting tasks).

Backs claims that HR does not fare well as a diagnostic indicator of workload. By employing a principal components analysis, it is possible to use the more or less standard measures of cardiovascular activity: heart rate or inversely heart period, the heart rate variability spectrum broken down into three frequency bandwidths thought to correspond to sources of autonomic activation, and residual heart period. The latter, RHP is usually a poor index of workload. The other measures have been shown to have reasonable value in detecting extremes in workload (eg., resting vs. work), as there is some evidence for diagnosticity, especially for HP and HR and occasionally, mid-band HRV. The PCA generally produces one factor associated with parasympathetic activity. The most consistent findings indicate that the four variables load on two factors, typically accounting for approximately 50% and 30% of the variance. The first factor is associated with parasympathetic activity and loads mid-band HRV and RSA, while the second factor is associated with sympathetic activity and loads HP and Residual HP. The factor loadings of these four variables are used to produce parasympathetic and sympathetic component scores, which are then subjected to the same analyses used for the original variables. To the extent that these composite scores produce more consistent outcomes, they will be valuable as diagnostic tools.

Cardiovascular activity in quasi-operational tasks. Rau (1996) used simulations of an electrical distribution system (electroenergy network) with trained operators. Two operators worked during each scenario, one as the shift leader and the other as a co-operator. Three types of tasks performed during system operation were chosen to reflect different levels of cognitive workload. Comparisons were made among these three workload conditions using HR. Heart rate was lower for the least demanding condition

and increased during the more demanding conditions, which were not different. Also, the shift leader showed higher HR during the most demanding task than the co-operator.

Veltman and Gaillard (1996) analyzed IBI and mid and high band HRV from subjects working in a flight simulator. A secondary CMT was included to increase cognitive workload. For analysis, the flight scenario was divided into five segments: rest periods, flight, flight with CMT, landing, post landing. IBI was longer (slower HR) during the rest periods than all flight segments, but no effect was seen for HRV bands. A comparison among the four remaining "flight" segments found that IBI was shorter (faster HR) for the flight with CMT and landing segments than for flight alone (diagnostic), while HRV in both bands was lower and equal for the three flight segments, than during the post-landing segment, which showed greater variablility. Veltman and Gaillard (1998) used pilots in a flight simulator with a flight scenario with 4 levels of maneuvering/pursuit difficulty. They measured heart period IBI and mid- and high-band HRV. The IBI was longer and HRVs were greater during a resting baseline than all flight segments. Comparisons among the levels of task difficulty found that IBI was diagnostic, with IBI decreasing (faster HR) as the task difficulty increased. HRV was not sensitive to task differences.

Tattersall and Hockey (1995) examined flight engineers in a flight simulator using HR and the mid- and high-bands of the HRV spectrum. The flight phase was divided into the takeoff/landing segment, and three levels of cognitive task demands during the cruising segment: system monitoring, routine fault correction, and problem solving. Compared to a baseline condition, HR increased and HRVs decreased during flight segments. During the flight segments, HR was higher during takeoff/landing than the in-flight cognitive tasks, which were not different. For, HRV only the mid-band was significant, with more suppression of variability for the demanding problem solving tasks, that the other two task types.

Backs, et al. (1999) used pilots in a Boeing 747 simulator with low and high workload scenarios. Five segments of the two flight scenarios (takeoff, top of climb, cruise, approach, and landing) were analyzed. Four cardiovascular measures were derived: Heart Period (interbeat interval), mid band HRV, high band HRV or Respiratory Sinus Arrhythmia (RSA), and Residual Heart Period. RHP is the heart period that remains after removing RSA, resulting in an index of sympathetic input to the heart. This measure is related to Residual Heart Rate, which removes the linearly related effect of respiratory activity on heart rate (Mulder, et al., 2000). A principal components analysis of these four variables estimated the relative contribution of the parasympathetic and sympathetic nervous systems and produced scores for each component. Importantly, the authors present reliabilities for each of the 6 measures in this design and HP was clearly the only statistically and clinically reliable measure. HP was shorter (faster HR) for the high workload scenario. Additionally, HP increased (slower HR) from takeoff to the cruise segment. HRV changes across flight segments are consistent with HP with suppression of HRV with higher workloads.

Overall, the research on heart-rate measures suggest that the mid-band HRV can accurately measure changes in mental workload and retains the properties of diagnosticity, sensitivity, reliability, and ease of use. Therefore, HRV has the potential, like EEG and ERPs, to be used as a physiological "trigger" for invoking adaptive automation.

Research Purpose

The EEG, ERP, and HRV represent viable candidates for determining shifts between modes of automation in adaptive systems. Because real-time assessment of workload is the goal of system designers wanting to implement adaptive automation, it is likely that these measures will become the focus of research on adaptive automation. This optimism stems from a number of studies that have suggested that they might be useful for on-line evaluations of operator workload (Defayolle et al., 1971; Farwell &

Donchin, 1988; Gomer, 1981; Humphrey & Kramer, 1994; Kramer, 1991; Kramer et al., 1989; Sem-Jacobsen, 1981). Although these results suggest that on-line assessment of mental workload may be possible in the near future, a good deal of additional research is needed.

Three experiments are reported that examined the efficacy of the EEG, ERPs, and HRV for adaptive task allocation. The studies are based on the pioneering research in physiological measures and adaptive automation reported by Pope, Bogart, and Bartolome (1995) who examined the use of EEG as an adaptive trigger for changing among automation task modes. They developed a biocybernetic system that has been validated in numerous studies to be capable of assessing candidate physiological measures for adaptive automation. Therefore, the three studies presented in the NASA Technical Paper utilized the experimental protocols of the biocybernetic system to assess the utility and potential of EEG, ERPs, and HRV for future human-centered adaptive automation design. The biocybernetic system is first described below followed by descriptions of Experiment 1 – 3, which examined the use of electroencephalogram, event-related potentials, and heart-rate variability for adaptive aided function allocation, respectively.

The Biocybernetic System

Pope, Bogart, and Bartolome (1995) reported one of the few studies examining the utility of EEG for adaptive automation technology. These researchers developed an adaptive system that uses a closed-loop method to adjust modes of automation based upon changes in the operator's EEG patterns. The closed-loop method was developed to determine optimal task allocation using an EEG-based index of engagement or arousal. The system uses a biocybernetic loop that is formed by changing levels of automation in response to changing taskload demands. These changes were made based upon an inverse relationship between the level of automation in the task set and the level of pilot workload.

The level of automation in a task set could be such that all, none, or a subset of the tasks could be automated. The task mix is modified in real time according to operator's level of engagement. The system assigns additional tasks to the operator when the EEG reflects a reduction in task set engagement. On the other hand, when the EEG indicates an increase in mental workload, a task or set of tasks may be automated, reducing the demands on the operator. Thus, the feedback system should eventually reach a steady-state condition, and neither sustained rises nor sustained declines in the EEG should be observed.

One issue for the biocybernetic system concerns the nature of the EEG signal used to drive changes in task mode. Pope, Bogart, and Bartolome (1995) argued that differences in task demand elicit different degrees of mental engagement that could be measured through the use of EEG-based engagement indices. These researchers tested several candidate indices of engagement derived from EEG power bands (alpha, beta, & theta). These indices of engagement were derived from recent research in vigilance and attention (Davidson, 1988; Davidson et al., 1990; Lubar, 1991; Offenloch & Zahner, 1990; Streitberg, Rohmel, Herrmann, & Kubicki, 1987). For example, Davidson et al. (1990) argued that alpha power and beta power are negatively correlated with each other to different levels of arousal. Therefore, these power bands can be coupled to provide an index of arousal. For example, Lubar (1991) found that the band ratio of beta/theta was able to discriminate between normal children and those with attention deficit disorder.Pope and his colleagues (1995) reasoned that the usefulness of a task engagement index would be determined by a demonstrated functional relationship between the candidate index and task operating modes (i.e., manual versus automatic) in the closed-loop configuration. They used both positive and negative feedback controls to test candidate indices of engagement because each should impact system functioning in the opposite way, and a good index should be able to discriminate between them. For example, under negative feedback conditions, the level of automation in the tasks was lowered (i.e., automated) when the EEG index reflected increasing engagement. On the other hand, when the EEG

reflected increases in task demands, automation levels were increased. Task changes were made in the opposite direction under positive feedback conditions; that is, the level of automation in the tasks was maintained when the EEG engagement index reflected increasing task demands. If there was a functional relationship between an index and task mode, the index should demonstrate stable short-cycle oscillation under negative feedback and longer and more variable periods of oscillation under positive feedback. The strength of the relationship would be reflected in the degree of contrast between the behavior of the index under the two feedback contingencies.

Pope, Bogart, and Bartolome (1995) found that the closed-loop system was capable of regulating participants' engagement levels based upon their EEG activity. They reported that the index 20 beta/(alpha+theta) possessed the best responsiveness for discriminating between the positive and negative feedback conditions. The conclusion was based upon the increased task allocations in the negative feedback condition witnessed under this index than under either the beta/alpha or alpha/alpha indexes. These results were taken to suggest that the closed-loop system provides a means for evaluating the use of psychophysiological measures for adapting automation. Recently, an improvement had been made to the biocybernetic system. The previous system used by Pope, Bogart, and Bartolome initiated changes in automation levels based on the slope of the index taken from successive measurements. One problem with using a slope measure concerns its sensitivity to changes in operator arousal and its reflection of levels of operator engagement. The system makes task allocation decisions regardless of whether the engagement level is high or low. In other words, an operator's overall engagement level may be quite low relative to his or her normal baseline engagement level. However, the system may make a task allocation decision to automate a task merely because the arousal level is higher, when the next EEG engagement index is derived, despite the fact that the overall arousal level is still low (Hadley, et al., 1997; Prinzel, Scerbo, Freeman, & Mikulka, 1997). Therefore, the system makes task allocation decisions without a consideration of individual differences in engagement.

Experiment One

Pope, Bogart, and Bartolome (1995) found that it was possible to moderate the level of engagement through a closed-loop system driven by the operator's EEG activity. Further, the index beta/(alpha + theta) showed the greatest difference between the positive and negative feedback conditions. There were more task allocations in the negative feedback condition than in the positive feedback condition with this index than with any of the other three indices. Moreover, they concluded that substituting either high beta (38-42 Hz) or EMG (42-100 Hz) in the numerator of the index would not significantly impact the ability of the beta/(alpha+theta) index to discriminate between feedback conditions.

Although the results of Pope, Bogart, and Bartolome (1995) show promise for designing adaptive automation technology around nonintrusive psychophysiological input, a number of limitations in their study must be addressed. Foremost, it remains to be seen whether this physiologically-based method of adaptive aiding can regulate performance, subjective workload, or task engagement, none of which were systematically examined in that study.

The present experiment, therefore, was designed to replicate and expand upon the original study by Pope, Bogart, and Bartolome (1995). We used a similar system to examine the effectiveness of the engagement index, beta/(alpha + theta), to produce expected feedback control behavior. Thus, the value of the index was expected to oscillate in a more regular and stable pattern under negative feedback than under positive feedback. Consequently, more task allocations were expected under the negative feedback than the positive feedback condition.

The results of Pope, Bogart, and Bartolome (1995) were generated with only a single compensatory tracking task. At present, however, it is not known how differences in task load would impact the manner in which tasks are allocated. Therefore, a second objective of the current study was to examine system operation under both single and multiple task conditions. Multiple resource theory (Wickens, 1984; 1992) posits that performance on a task that is performed in conjunction with other tasks should be poorer than performance on a task performed alone because of competition for cognitive resources. For example, when Parasuraman, Molloy, and Singh (1994) asked participants to perform either a system monitoring task (single task condition), or a system monitoring task, compensatory tracking task, and a resource management task (multiple task condition), they missed fewer critical signals while performing the system monitoring task alone than when performing all the tasks concurrently. Results such as these are not limited solely to monitoring tasks. For example, Arnegard (1991) found that the combination of these same three tasks resulted in a significant increase in workload compared to only the compensatory tracking task. The results of these studies suggest that multiple task conditions produce higher levels of workload and can lead to decreases in performance.

Automation-induced performance decrements in multiple task environments may stem from changes in the processing strategies that participants use to devote cognitive resources to the different tasks. A number of researchers have stated that operators may become complacent as they gain more experience with automation leading to an increase in trust and reliance on automation (Riley, 1994; Singh, Molloy, & Parasuraman, 1993). Such shifts in strategy do not provide adequate processing resources for the maintenance of automated tasks. It has been suggested that adaptive systems, however, are less susceptible to automation-induced performance decrements because of the regulation of workload and maintenance of operator engagement (Hancock & Chignell, 1988; Scerbo, 1996). The closed-loop system was designed to moderate workload by reducing task demands when levels of workload increase. Accordingly, we expected that the biocybernetic system would make more task allocations under the multiple task condition in order to compensate for the increased fluctuations in taskload that would accompany the operation of multiple tasks each with their own unique demand schedules. Furthermore, performance under the multiple task condition was predicted to be significantly better for participants who performed these tasks under the closed-loop system than a control group who performed these tasks without the benefit of adaptive task allocation.

Pope, Bogart, and Bartolome (1995) argued that a closed-loop feedback system between the pilot, the equipment being monitored, and physiological recording devices provides a means for maintaining optimal states of arousal and performance in a flight environment. However, they did not report any performance data to substantiate this notion. Furthermore, little research is available that examines the relationship between performance, mental workload, and physiological indices in such a biocybernetic system. Thus, a final objective was to verify system operation with both performance and physiological data as well as with subjective estimates of workload.

Method

Participants

Forty-eight participants were used for this experiment. The ages of the participants ranged from 18 to 40. Half of the participants had some flight training, but all had significant experience with flight simulation software.

Apparatus

Electrical cortical activity was recorded with an Electro-Cap International sensor cap. The lycra sensor cap consists of 22 recessed tin electrodes arranged according to the International 10-20 system (Jasper, 1958). One mastoid electrode was used for a reference. Conductive gel was placed into each of the four electrode sites, the reference, and the ground using a dispenser tube and a blunt-tipped hypodermic needle.

The EEG amplification system was a BIOPAC EEG100A differential amplifier module. The system consists of a four separate channel, high gain, differential input, bio-potential amplifier. The frequency response was 1 to 100 Hz. The gain was set at x5000 and allowed an input signal range of 4000*u*V (peak-to-peak).

The EEG100A was connected to a Macintosh Virtual Instrument (VI). The software designed to run the VI calculated the total EEG power in three bands: theta (4-8 Hz), alpha (8-13 Hz), and beta (13-22 Hz). The VI also performed the engagement index calculations and commanded the task mode changes through serial port connections to the task computer.

The Macintosh Virtual Instrument was connected to a WIN 386 SX computer with a NEC MultiSync 2A color monitor that was used to run the MAT. An Analog Edge joystick was used for the compensatory tracking task. The joystick was set to have a gain of 60% of its maximum.

Experimental Tasks

Participants operated a modified version of the NASA Multi-Attribute Task (MAT) Battery (fig. 1; Comstock & Arnegard, 1992). The MAT Battery is composed of four separate task areas or windows constituting the monitoring, compensatory tracking, communication, and resource management tasks. These different tasks were designed to simulate activities that airplane crew members often perform during flight. Only the monitoring, compensatory tracking, and resource management tasks were used for the present study. The functioning of the monitoring and resource management tasks were controlled by a script file that controlled the sequence and timing of the events in the tasks. The compensatory tracking task was cycled between manual and automatic modes at preset times for those participants in the control group. However, the amount of time that these participants spent controlling the tracking task in each of these task modes was approximately equal to the time spent by participants in the experimental group ($\underline{p} > .05$).

Experimental Design

A 2 feedback condition (positive or negative feedback) X 2 task mode (automatic or manual mode) X 2 task level (single or multiple task condition) X 2 group (experimental or control group) mixed-subjects design was employed. The group condition represented the only nested variable. All other experimental conditions were counterbalanced. The dependent variables were EEG engagement index as well as the relative power of theta, beta, and alpha at each cortical site (see below). Another dependent variable was the number of switches, or task allocations, under each feedback condition. Performance was measured by root-mean-squared-error (RMSE) and subjective workload was assessed by the NASA-TLX (Hart & Staveland, 1988).

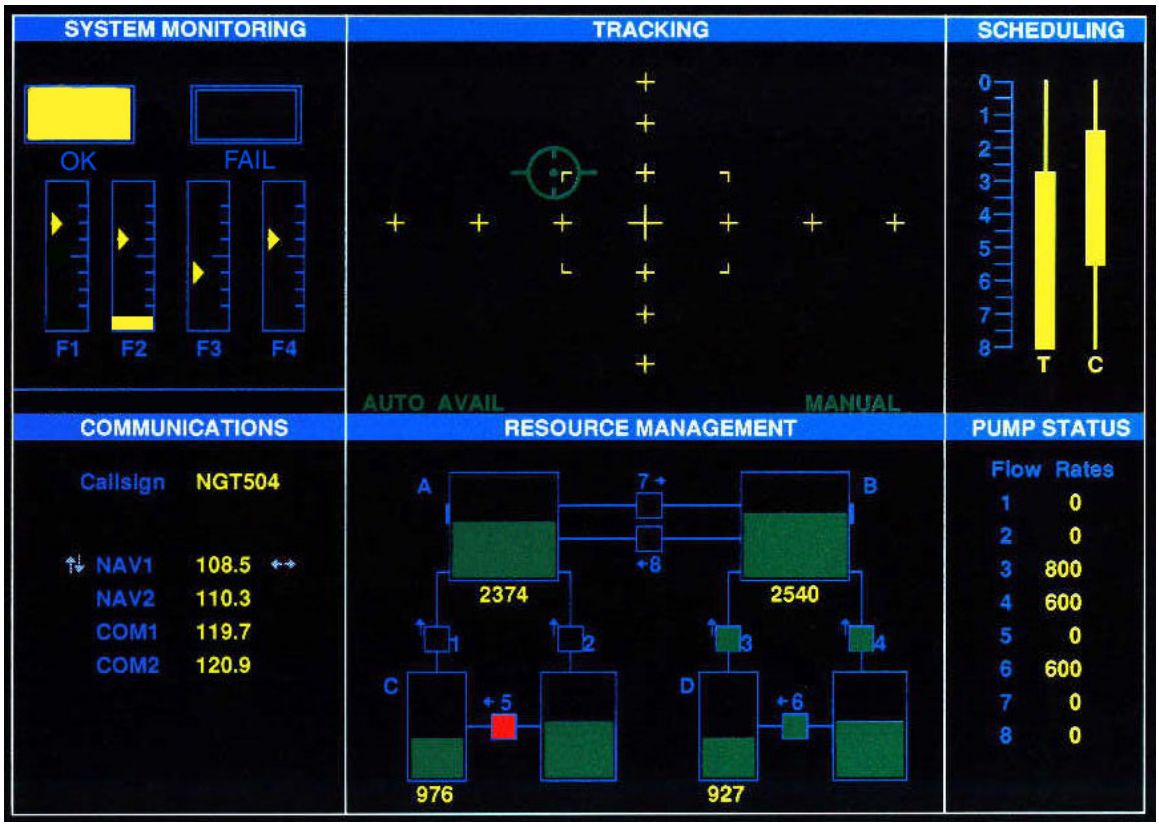


Figure 1. The Multi-Attribute Task Battery (printed in grayscale).

EEG Recording and Analysis

The EEG was recorded from sites Pz, Cz, P3, and P4. A ground site was located midway between Fpz and Fz. Each site was referenced to the left mastoid. Each amplified EEG channel was digitized at a rate of 400 samples per second. The digital signals were arranged into epochs of 1024 data points (roughly 2.5 seconds) prior to conversion to a spectral power form using a Fast Fourier Transform (FFT). Digitized input channels were converted back to analog then routed to an EEG interface with a LabVIEW Virtual Instrument (VI). The VI calculated total EEG power from the bands of theta, alpha, and beta for each of the four sites. The EEG frequency bands were set as follows: alpha (8-13 Hz), beta (13-22 Hz), and theta (4-8 Hz). The VI also calculated the EEG engagement index that determined the MAT Battery task mode changes. The beta / (alpha + theta) index was used in the present study because it was shown to be the most sensitive by Pope, Bogart, and Bartolome (1995).

Task mode was switched to either manual or automatic depending upon the feedback condition. The index was calculated every 2 sec with a moving 40-sec window procedure. The slope between successive calculations was then determined. An increasing slope represented increasing task engagement and a decreasing slope represented decreasing task engagement. An artifact rejection subroutine examined the amplitudes of each epoch from the four channels of digitized EEG and compared them with a preset threshold. If the voltage in any channel exceeded the threshold for more than 25% of the epoch (about two-thirds of a second) the epoch was marked as an artifact and the calculated index was replaced with a value of zero. These epochs were then ignored when computing the average value of the index. The data record resulting from an epoch containing an artifact was marked when it was written to the data file so

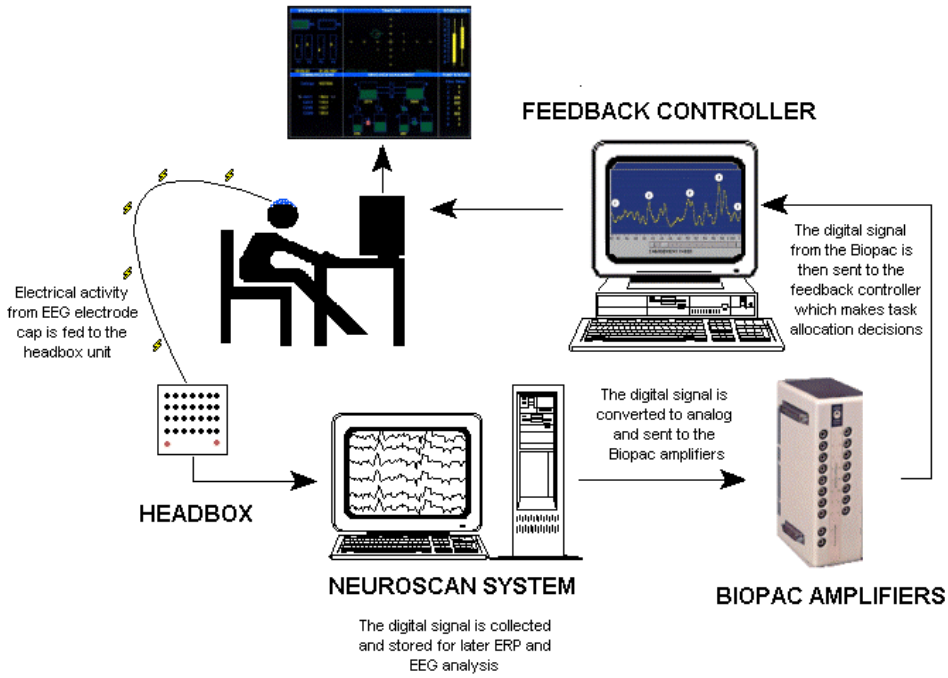


Figure 2. The Closed-Loop System.

that it could be ignored during later data analyses. Figure 2 shows a graphical depiction of the closed-loop system.

Experimental Procedure

The participant's scalp was prepared with rubbing alcohol and electrolyte gel. A reference electrode was then affixed to the participants left mastoid by means of electrode tape and an adhesive pad. Electrode gel was then placed in each of the four electrode sites (Pz, Cz, P3, P4), the ground site, and the reference electrode with a blunt-tip hypodermic needle. The scalp was lightly abraded to reduce the impedance level of the sites, relative to the ground, to less than five KOhms as measured by a Nether Electrode Impedance Meter. The participant was then brought into the experimental room and hooked up to the BIOPAC EEG100A amplifier. Participants in the control group, however, were not fitted with the EEG electrode cap.

Participants in both the experimental and control groups were given 25 minutes of practice with the MAT Battery. This was done to ensure that learning effects would not confound task performance over experimental trials. The practice time of 25 minutes was determined by a pilot study of 24 participants who performed all three tasks simultaneously for 30 minutes. The results of the pilot study revealed that participants did not improve after 25 minutes and this was confirmed by self-report measures of performance. After practice, participants were given five minutes of rest before the actual data collection took place. Participants in the experimental group were asked to perform in both a single and multiple task condition each lasting 16 minutes. Those in the control group performed only the multiple task condition also for a period of 16 minutes. After each task run, participants were asked to fill out the NASA-TLX. After completing all experimental trials, participants were debriefed.

Results

All ANOVAs using a repeated measures variable were corrected with the Greenhouse-Geisser procedure (Greenhouse & Geisser, 1959). Alpha level was set at .05 and all post hoc comparisons were computed using simple effects analyses and the Tukey post hoc procedure.

Task Allocations

A main effect was found for Task Level, $F(1,23) = 5.09$. There were significantly more task allocations under the multiple task condition ($M = 37.67$) than under the single task condition ($M = 34.08$). There was also a significant main effect for feedback condition, $F(1,23) = 4.29$. More task allocations were made under negative feedback ($M = 37.20$) than under positive feedback ($M = 33.95$).

No differences were found in time spent in automated and manual task modes ($p > .05$). Participants spent approximately the same amount of time performing the MAT monitoring and resource management tasks under both automated and manual tracking modes.

Electroencephalogram

A MANOVA, performed on the EEG engagement index and the three relative power bands at the four cortical sites, revealed a significant interaction of Feedback Condition and Task Mode, $F(14,96) = 18.16$. Subsequent ANOVAs showed a significant interaction of Feedback Condition and Task Mode for the EEG engagement index, $F(1,23) = 145.22$; as well as the theta band, $F(1,23) = 33.04$; alpha band, $F(1,23) = 29.34$; and beta band, $F(1,23) = 76.42$. Table 1 presents the means and standard deviations for the Feedback Condition X Task Mode interaction.

RMSE

A significant main effect for performance on the compensatory tracking task was found for Task Level, $F(1,23) = 78.57$. Tracking error was significantly lower in the single task condition ($M = 8.90$) than in the multiple task condition ($M = 15.22$). Participants also performed significantly better in the negative feedback condition ($M = 5.84$) than in the positive feedback condition ($M = 11.25$), $F(1,23) = 6.67$. It is important to note that RMSE was analyzed only for tracking performance in the manual task mode.

An analysis examining tracking performance under the multiple task condition between those participants operating in the biocybernetic system and the control group revealed a main effect, $F(1,47) = 4.049$. Participants in the experimental group performed significantly better ($M = 15.22$) than the participants in the control group ($M = 17.90$).

NASA-TLX

The total TLX score was found to be significant for Task Condition, $F(1,23) = 46.05$. Thus, participants rated the multiple task condition ($M = 76.2$) to be significantly higher in subjective workload than the single task condition ($M = 36.4$). There was also a main effect for total TLX score between the experimental group ($M = 76.2$) and control group, ($M = 92.5$) $F(1,47) = 5.105$.

Discussion

Experiment One reported on a closed-loop, biocybernetic system developed to test various psychophysiological measures for their use in adaptive automation. Specifically, we assessed the use of the EEG band ratio, beta/(alpha+theta) on the basis of behavioral, system, and physiological data gathered under negative and positive feedback controls. Furthermore, the study was designed to determine how different taskloads impact adaptive task allocation and system regulation of task engagement and workload.

Task Allocations

Regarding task allocations, Pope, Bogart, and Bartolome (1995) stated that the relative usefulness of a task engagement index can be found in the relationship between it and automation mode under negative and positive feedback controls. Specifically, they evaluated the ability of the index to produce expected differences in system operation between positive and negative feedback. Under negative feedback, loss of engagement should trigger increased task demand that results in a task allocation to a manual-operating mode. However, a loss of engagement under positive feedback would instead result in a task allocation to (or maintenance in) the automatic operating mode. Therefore, the system should oscillate more frequently under the negative feedback condition in order to maintain a stable level of engagement. On the contrary, positive feedback should produce longer episodes in each of the task modes and, consequently, there should be fewer task allocations.

Pope, Bogart, & Bartolome (1995) reported that three indices, beta/alpha, beta/(alpha+theta), and alpha/alpha were able to distinguish between the feedback conditions, but the best discriminator was the index, beta/(alpha+theta). This index resulted in significantly more task allocations under the negative feedback condition than under the positive feedback condition. The results of the present study also show that there were more task allocations under the negative feedback condition with the beta/(alpha+theta) index. In addition, other studies (Hadley, Mikulka, Freeman, Scerbo, & Prinzel, 1997; Prinzel, Freeman, Scerbo, & Mikulka, 1997; Prinzel, Scerbo, Freeman, & Mikulka, 1997) have also shown that this index best produces expected feedback control behavior.

Electroencephalogram

Pope, Bogart, and Bartolome (1995) argued that the closed-loop feedback system provided a method for regulating operator attention, arousal, and workload. However, the only evidence that these researchers reported was the number of task allocations made between negative and positive feedback. Because the closed-loop system is based upon theories that relate EEG to levels of operator workload, it seems that a more valid measure would be the actual value of the EEG engagement index.

The interaction between feedback condition and task mode for the EEG engagement index and power bands provides validation that these psychophysiological measures are responsible for the operation of the system. Under positive feedback, when EEG patterns reflected high task engagement, characterized by increased beta, alpha blocking, and theta suppression, the tracking task was set to the manual task mode. However, when the EEG patterns reflected low task engagement, the system automated the tracking task. Therefore, the value of the EEG engagement index was expected to be largest under the positive feedback, manual task mode and smallest under the positive feedback, automatic task mode. The EEG engagement index did indeed show this pattern.

On the contrary, negative feedback was expected to produce an opposite pattern of results. The EEG engagement index was predicted to be higher under the negative feedback, automatic task mode because

the system assigns the tracking task to the operator when the EEG engagement index reflects decreasing task engagement and maintains this task mode as long as the value of the index remains low. Again, the results demonstrate that the system operated accordingly.

The pattern of results for the EEG engagement index reflect characteristics of the theta, alpha, and beta power bands that comprise the index. Research has shown that increases in arousal, attention, and workload are followed by decreases in theta and alpha power and an increase in beta power (Davidson, 1988; Kramer, 1991; Parasuraman, 1983; 1984; Sterman & Mann, 1995). Therefore, one should expect that theta and alpha should be highest under the positive feedback, automatic task mode and the negative feedback, manual task mode, but beta should be highest under the positive feedback, manual task mode and negative feedback, automatic task mode. The feedback condition by task mode interactions for each of the three EEG power bands indicate that theta, alpha, and beta each contributed to system operation under the different operating task modes and feedback conditions.

The results from the individual power bands may suggest that theta, alpha, or beta alone are capable of driving the closed-loop system as effectively as the beta/(alpha+theta) engagement index. Although a comparison of these individual power bands has not yet been undertaken with the closed-loop system, the results of Pope, Bogart, and Bartolome (1995) argue against such a conclusion. These researchers found that the alpha power band alone was not as reliable an index as the beta/(alpha+theta) index. Prinzel et al. (1997) also found that the index, 1/alpha, did not distinguish between positive and negative feedback conditions. Moreover, a recent study focusing on the contribution of theta, alpha, and beta have in the operation of the closed-loop system suggest that it is the combination of these three power bands that produces the strongest outcomes rather than any individual power band (Freeman, Clouatre, Pickett, Mikulka, & Scerbo, 1995).

Tracking Performance

In the original study by Pope and his colleagues (1995), only the number of task allocations was reported to show the efficacy of the system for regulating operator engagement. Even if task allocations were greater under negative feedback, there is little practical value of such a system if it does not also have an impact on performance. As Byrne and Parasurman (1996) stated, the effects of various interventions, such as changes in task allocations, should be assessed using other workload measures and tools. Furthermore, they noted that any assessment of the use of psychophysiological measures for adaptive automation must be made in conjunction with measures of performance. Accordingly, performance under both negative and positive feedback was analyzed in the present study. As predicted, tracking performance was found to be significantly better under the negative feedback condition than under the positive feedback condition. These results suggest that the closed-loop system can facilitate performance and compliments the task allocation and psychophysiological data supporting the use of the system for adaptive task allocation.

Task Load

Although one of the goals of the closed-loop system described by Pope, Bogart, and Bartolome (1995) was to moderate workload, they did not report any measures of workload. This issue was examined directly in the present study by including a single and a multiple task condition representing low and high workload conditions, respectively. It was predicted that the closed-loop system would make more task allocations under the high workload condition because of the unpredictable workload demands associated with the performance of the three different tasks. In addition, workload ratings and tracking error were expected to be higher under the multiple task condition.

The results showed that more task allocations were made under the multiple task condition. Therefore, the system appears to be sensitive to increases in taskload. Participants also rated workload higher and performed the tracking task more poorly under the high workload condition. The EEG engagement index, however, was not found to discriminate between these two task conditions although the value of the index was higher under the multiple task condition than under the single task condition. Nevertheless, these results support that the single and multiple task conditions provided different levels of taskload.

If the system does indeed moderate workload and task engagement, there should be a significant reduction in workload and performance errors while operating the tasks in the closed-loop environment. Therefore, we compared performance between an experimental group, who performed the three tasks under the closed-loop system, with a control group who performed these same three tasks without the closed-loop system. Again, the amount of time spent in the manual task mode was comparable between the two groups. The results showed a significant difference in performance and workload between the two groups. Participants in the experimental group rated workload lower and had lower tracking error scores than those participants in the control group. Therefore, these results suggest that the closed-loop system is capable of moderating operator workload and improving performance through an adaptive system driven by an operator's EEG patterns.

Experiment Two

The results of Experiment One suggest that the closed-loop system represents a method for the use of psychophysiological measures in adaptive automation technology. However, because the closed-loop system has only been used for testing EEG indices, it remains to be seen whether other psychophysiological measures will also be appropriate for use with this system. Therefore, Experiment Two was designed to examine the efficacy of event-related potentials.

Experiment Two attempted to further the research on the use of ERPs for adaptive automation. The same biocybernetic system was used to make task allocation decisions between manual and automatic task modes as previously described. Participants were also asked to perform an oddball, auditory task concurrently with the compensatory tracking task. The EEG signal was fed to both the biocybernetic system and to a data acquisition system that permitted the analysis of ERPs to high and low frequency tones. It was hypothesized that the amplitude of the ERP components would be higher and latency shorter for events elicited in the secondary task under the adaptive automation condition compared to either a yoked or control group condition.

Method

Participants

Thirty-six subjects participated in the experiment. The ages of the participants ranged from 18 to 40. All participants were right-handed as measured by the Edinburgh handedness survey (Oldfield, 1971) and had normal or corrected-to-normal vision. Twelve of the participants had some flight training, but all thirty-six reported "substantial" experience with flight simulation software and were pre-screen for proficiency before selection for research.

Apparatus

Electrical cortical activity was recorded with an Electro-Cap International sensor cap. The lycra sensor cap consists of 22 recessed tin electrodes arranged according to the International 10-20 system

(Jasper, 1958). One mastoid electrode was used for a reference. Conductive gel was placed into each of the four electrode sites, the reference, and the ground using a dispenser tube and a blunt-tipped hypodermic needle.

The NeuroScan SynAmps is a AC/DC amplifier that provides both a broadband amplifier and a high speed digital acquisition system. The system has four high-speed digital signal processors (DSPs) with 1 MByte of RAM per DSPs for data acquisition. The SynAmps has a 33 MHz 486 DX processor with 4 MBytes of RAM and an electronic flash disk dedicated to management of DSPs. It provides for real-time digital filtering by the DSPs allowing filter settings from DC to 10kHz. Sampling rates can be set between 100 Hz to 20 kHz from 1 to 32 channels. Also, the system has 28 monopolar and 4 bipolar channels provided through a NeuroScan SynAmps headbox connector. The SynAmps amplifier has tracking anti-aliasing filters, first stage amplification to reduce Signal/Noise ratio, and an on-line DC offset correction. All impedance calibration is built-in and the input signal is managed through SCAN software. The system was used for ERP acquisition and analyses.

The SynAmps amplifier was connected via an analog output board to a Biopac EEG100A Analog/Digital converter through a four-line buffered cable. The analog output board takes the output signal from the SynAmps prior to the sample and hold (S/H) circuits. The analog output board filters the signal and then routes the output to a D-37 connector on the SynAmps back panel. Band-limiting is gathered from single-pole high-pass (1 Hz) and low-pass (70 Hz) filters. The anti-aliasing filters are set for 0.2 times the sample frequency.

The system was also connected to a PC computer through the parallel port on the back panel of the SynAmps amplifier. The Biopac system consists of a four channel, high gain, differential input, bio-potential amplifier. The frequency response is 1 to 100 Hz. The gain setting is x5000 that allows an input signal range of 4000*u*V (peak-to-peak). However, for the present study, only the Biopac A/D converter was used.

The Biopac A/D converter was connected to the Macintosh Virtual Instrument (VI). The software designed to run the VI is the Real Time Cognitive Load Evaluation System (RCLES v 3.3.1). It calculates the total EEG power in four bands: theta (4-8 Hz), alpha (8-13 Hz), beta (13-22 Hz), and high beta (38-42 Hz). The VI also performs the engagement index calculations and commands the task mode changes through serial port connections to the task computer.

The Macintosh Virtual Instrument was connected to a PC WIN 486 DX computer that was used to run the MAT (see below). Data was binned according to assigned bit numbers placed in the data record from the PC computer. Auditory oddball tone sequencing and gating was controlled by the VI software and these event signals were also placed in the data record as ERP synchronization triggers.

The monitor was a NEC MultiSync 2A color monitor. A joystick was used for the compensatory tracking task. The gain on the joystick was set to 60% of its maximum and had a bandwidth of 0.8 Hz. A graphical depiction of the experimental set-up is shown in figure 1.

Experimental Design

A 2 feedback condition (positive or negative feedback) X 2 task mode (automatic or manual mode) X 3 experimental group condition (yoked, control, or adaptive automation) mixed-subjects design was employed. The experimental group condition represented the only nested variable. All other conditions were counterbalanced.

Automation Cycle Sequencing. Each of the thirty-six participants was randomly assigned either to the adaptive automation group (n = 12), the yoked (n = 12), or the control (n = 12) group. The adaptive automation condition required the participants to perform the compensatory tracking task and auditory oddball task under the closed-loop configuration. The data records of switches between task modes were then used to determine the pattern of task allocations to be made between automatic and manual task modes for participants in the yoked condition. Therefore, these participants performed the tracking task under the exact same schedule of manual and automatic task modes as their experimental complement. The control group, on the other hand, consisted of participants who performed a random assignment of task allocations between task modes. The schedule of task allocations was determined for each control participant based upon the average number of switches in both the positive and negative feedback conditions for the adaptive automation group. For example, control participant number one received a random schedule of task allocations based upon the average number of task allocations that adaptive automation participant number one experienced. All participants, however, had the same sequence of high and low tones in the auditory oddball task.

Dependent Variables. The dependent variables included: (a) the EEG engagement index defined as 20 beta / (alpha+theta); (b) the amplitude and latency of the ERP waveform was analyzed; (c) the number of switches, or task allocations, under each feedback condition; (d) tracking performance as measured by root-mean-squared-error (RMSE); (e) the number of counted high tones in the oddball task; and (7) subjective workload assessed by the NASA-TLX (task load index; Hart & Staveland, 1988; Byers, Bittner, & Hill, 1989).

Statistical Tests and Criterion. All ANOVAs using a repeated measures variable were corrected with the Greenhouse-Geisser procedure (Greenhouse & Geisser, 1959). Alpha level was set at .05. All post hoc comparisons used simple effects analyses and the Tukey post hoc procedure.

Experimental Tasks

Tracking Task. Participants were run using a modified version of the NASA Multi-Attribute Task (MAT) battery (Comstock & Arnegard, 1992). The MAT battery is composed of four separate task areas, or windows, constituting the monitoring, compensatory tracking, communication, and resource management tasks. These different tasks were designed to simulate the tasks that airplane crewmembers often perform during flight. Only the compensatory tracking task was used in the present study. The task requires participants to use a joystick to maintain a moving circle, approximately 1 cm in diameter, centered on a .5 cm by .5 cm cross located in the center of the screen. Failure to control the circle results in its drifting away from the center cross.

Auditory Oddball Task. The auditory oddball secondary task consisted of high and low tones at 1100 Hz and 900 Hz, respectively. The frequency of the tone presentation was once per second, and was randomly assigned for presentation. The inter-stimulus interval was kept uniform across the experimental conditions. Therefore, over a 16-minute trial there were 96 high tone signals and 864 low tone signals. The ordering of the onset of tones was held consistent across participants. The tones were gated to provide a rise and fall time of .10 shaping a square wave signal. The tones were presented to both of the participant's ears through stereo KOSS head phones at 60 dB SPL.

EEG Recording and Analysis

The EEG was recorded from sites Pz, Cz, P3, and P4. A ground site was located midway between Fpz and Fz. Each site was referenced to the left mastoid. The EEG was routed through a SynAmps amplifier

from an analog output board to the Biopac A/D converter. The outputed analog signal was converted to digital by the BioPac A/D converter, and the digital signals were arranged into epochs of 1024 data points (roughly two and one half seconds). Digitized input channels were then converted back to analog and then routed to an EEG interface with a LabVIEW Virtual Instrument (VI). The VI calculated total EEG power from the bands of theta, alpha, and beta for each of the four sites and converted the signal into a spectral power form using a Fast Fourier Transform (FFT).

The EEG frequency bands were set as follows: alpha (8-13 Hz), beta (13-22 Hz), theta (4-8 Hz), and high beta (38-42 Hz). The VI also calculated the EEG engagement index that determines the MAT Battery task mode changes. Automation task mode was switched between manual and automatic depending upon the feedback condition. The EEG index was calculated every 2 sec with a moving 20-sec window. The window was then advanced two seconds and a new average was derived. This moving window process continued for the duration of the trial. At each epoch, the index was compared to the mean value determined during a five-minute baseline period for each participant. An EEG index above baseline (see below) indicated that the participant's engagement level was high while an EEG index below baseline indicated that engagement level was low. An artifact rejection subroutine examined the amplitudes of each epoch from the four channels of digitized EEG and compared them with a preset threshold. If the voltage in any channel exceeded the threshold for more than 25% of the epoch (about two-thirds of a second) the epoch was marked as artifact and the calculated index was replaced with a value of zero. These epochs were then ignored when computing the value of the index. The data record resulting from an epoch containing an artifact was marked when it was written to the data file so that it could be ignored during later data analyses.

ERP Recording and Analyses

The NeuroScan SynAmps amplifier system was used for ERP acquisition and analyses. The software package for gathering ERPs was the Acquire386 SCAN software version 3.00. Data was acquired based upon assigned bit numbers placed in the data record from the MAT computer. The signal was gathered with 500 sweeps and points in the time domain providing an A/D rate of 500. All corrections and artifactual rejection were done off-line. The amplifier had a gain setting of 500 with a range of 11 mV and an accuracy rate of 0.168 uV/bit. The low pass filter was 30 Hz and the high pass filter was set at 1.0 Hz. EEG electrodes had an impedance of below 5 KOhms.

The continuous EEG data file was analyzed to reduce ocular artifact through VEOG and HEOG electrodes. These channels were assigned weights according to a sweep duration of 40 ms and minimum sweep criteria of 20. The continuous EEG data file then transformed into an EEG epoch file based on a setting of 500 points per data file. The epoch file was then baseline corrected in the range of −100 to 0 msec from the onset of the signal. ERPs were acquired through a sorting procedure based upon the assigned bit numbers in the data file. The signal was then further filtered with a low pass frequency of 62.5 and a low pass slope of 24 db/oct. The high pass frequency was 5.00 Hz with a high pass slope of 24 db/oct. All filtering was performed in the time domain. All EEG was referenced to a common average and was smoothed by the SCAN software.

The criteria for ERP component classification was determined by the largest base-peak amplitude and latency within a pre-set window (Kramer, Trejo, & Humphrey, 1996): N100 (0-150 msec), N200 (150-250 msec), P100 (0-150 msec), P200 (150-250 msec), and P300 (275-750 msec).

Experimental Procedure

The participant's scalp was prepared with rubbing alcohol and electrolyte gel. A reference electrode was then affixed to the participant's left mastoid by means of electrode tape and an adhesive pad. ECI Electro-Gel conductive gel was then placed in the reference electrode with a blunt-tip hypodermic needle. Electrode gel was also placed in each of the four electrode sites (Pz, Cz, P3, P4), the ground site, and VEOG and HEOG electrodes. Using the blunt-tip hypodermic needle, the scalp was lightly abraded to reduce the impedance level at each site, relative to the ground, to less than five KOhms.

Participants were then instructed on how to perform the auditory oddball task and the compensatory tracking task. Once the participant had an understanding of these tasks, the EEG electrode cap was connected to the SynAmps headbox connector. Participants were then asked to sit quietly with their eyes open and then with their eyes closed for five minutes each. EEG was gathered during this time to establish baseline parameters. The mean EEG value during this time represented the baseline criteria for determining task allocations during the experimental session.

After gathering baseline data, participants were given a five-minute break and, thereafter, the experimental session began. For participants in the adaptive automation group, there were two experimental trials consisting of 16 minutes of either positive or negative feedback. Participants in the yoked and control conditions also had two 16-minute trials. However, the yoked participants performed the tasks based upon the schedule of task allocations of their yoked counterparts. For the control group, the two 16-minute trials consisted of a random assignment of the same number of task allocations between manual and automatic task modes for both positive and negative feedback that participants in the adaptive automation group experienced (see above).

After each experimental trial, all participants were asked to fill out the NASA-TLX (Hart & Staveland, 1988). After the experimental session is completed, all participants were debriefed.

Results

The data from the study were analyzed using a series of MANOVAs (multivariate analysis of variance) and ANOVAs (analysis of variance) statistical procedures. In all cases, alpha level was set at .05 and was used to determine statistical significance. The Greenhouse-Geisser procedure was used to correct psychophysiological data (Greenhouse & Geisser, 1971). Analyses of simple effects and Student Newman-Keuls (SNK) post-hoc tests were used to examine significant main and interaction effects.

Task Allocations

A simple ANOVA procedure was performed on the task allocation data for feedback condition for the adaptive group only. The negative feedback condition ($\underline{M}$ = 68.92) produced more task allocations than the positive feedback condition ($\underline{M}$ = 50.83), $F(1, 11) = 6.50$. An ANOVA also revealed that the amounts of time participants performed the tracking task in the automatic and manual task modes was not significantly different regardless of feedback condition, $F(1, 11) = 0.97$.

Tracking Performance

A 3 (group) X 2 (feedback) ANOVA revealed significant main effects for feedback condition, $F(1, 33) = 9.01$; and group condition, $F(2, 33) = 3.31$. Participants performed significantly better under the negative feedback condition (M = 8.91) than under the positive feedback condition (M = 11.14). Additionally, participants in the adaptive automation group did significantly better on the tracking task

(M = 8.55) than those participants in the yoked condition (M = 11.06) or in the control condition (M = 10.45).

There was also a group X feedback condition interaction for tracking performance, F (2, 33) = 4.84. Participants in the adaptive automation group had significantly lower tracking error when performing the task under the negative feedback condition than under any of the other group, feedback condition combinations.

Subjective Workload

A significant main effect was found for feedback condition, F (1, 11) = 39.83. Participants in the adaptive automation group rated the negative feedback condition to be lower in workload (M = 72.50) than the positive feedback condition (M = 87.66). There was also a main effect for group condition, F (2, 33) = 13.76. Those participants in the adaptive automation group reported overall workload (M = 63.70) to be much lower than those participants in the yoked condition (M = 88.04) or in the control condition (M = 88.50).

A group X feedback condition interaction was also found, F (2, 33) = 27.67. A simple effects analysis showed that participants in the adaptive automation group rated the negative feedback to be much lower in workload than under any of the other group, feedback condition combinations. No other differences were found to be significant.

Auditory Oddball Task Performance

There was a significant group X feedback condition interaction for secondary task performance, F (2,33) = 4.12. Participants, in the adaptive automation group, were more accurate in counting the number of high tones presented when they performed the task under the negative feedback condition (M = 94.32) than under the positive feedback condition (M = 83.29). Also, performance under the adaptive automation, negative feedback condition was significantly better than performance under the yoked group condition for positive feedback (M = 85.32) or negative feedback (M = 87.32). Additionally, performance for participants in the control condition for positive feedback (M = 84.32) or negative feedback (M = 84.98) was significantly poorer than when performing the task under the adaptive automation, negative feedback condition. Simple effects analyses found no differences between the yoked group or control group conditions. Furthermore, performance was not significantly different between these two group conditions and the adaptive automation, positive feedback condition.

Electroencephalogram

An ANOVA on the EEG engagement index for the adaptive automation condition revealed no main effects for feedback condition, F (1,11) = 0.89; or task mode, F (1,11) = 0.34. There was, however, a significant feedback condition X task mode interaction for the EEG engagement index, F (1, 11) = 201.32. A simple effects analysis found that the EEG engagement was higher during positive feedback, manual task mode (M = 11.91) and lower during negative feedback, manual task mode (M = 8.23). Also, the EEG engagement index was larger under the negative feedback, automatic task mode (M = 11.45) than under the positive feedback, automatic task mode (M = 8.10). No differences were found between the negative feedback, automatic task mode and the positive feedback, manual task mode. Additionally, there were no differences found between the negative feedback, manual task mode and the positive feedback, automatic task mode. Table 1 presents the mean values of the EEG engagement index.

Table 1. Means for EEG Engagement Index

	Task Mode	
	Manual	Automatic
Negative Feedback	8.12	11.83
Positive Feedback	11.98	8.05

Event-Related Potentials

Wilk's Lambda MANOVAs were performed on the base-peak amplitude and latency data for N100, P200, and P300 ERP components for electrodes Cz, Pz, P3, and P4. There were no significant effects found across the four electrodes, F (3, 33) = 1.12. Therefore, subsequent analyses were on collapsed data across electrode sites.

Significant effects were found for feedback condition, F (6, 28) = 13.64; group condition, F (12, 56) = 6.29; and group X feedback condition, F (12, 56) = 8.31. Therefore, subsequent ANOVAs were performed on these main effects and interaction for both ERP amplitude and latency.

N100 Amplitude. There was a significant main effect found for feedback condition, F (1, 11) = 4.93. The N100 amplitude tended to be larger under the positive feedback condition (*M* = -4.47) than under the negative feedback condition (*M* = -3.38). There was also a main effect found for group condition, F (2, 33) = 17.58. A Tukey post hoc test revealed that the amplitude was larger for those participants in the adaptive automation group (*M* = -4.49) and yoked group (*M* = -4.15) than in the control group (*M* = -3.15).

In addition to main effects, there was a group X feedback condition interaction, F (2, 33) = 13.00. N100 amplitude was significantly larger under the adaptive automation, negative feedback condition than under any other group X feedback conditions (See Tables 7-8). Simple effects analyses revealed no other significant effects for this interaction. The group X feedback condition interaction is presented in Table 2. Figure 3 presents the ERP graphically for the negative feedback contingency across groups.

N100 Latency. No main effects or interactions were found for feedback condition, F (1, 11) = 0.67; group condition, F (2, 33) = 0.94; or the group X feedback condition interaction, F (2, 33) = 0.79.

P200 Amplitude. No effects were found for feedback condition, F (1, 11) = 0.01; group condition, F (2, 33) = 2.87; or the group X feedback condition interaction, F (2, 33) = 0.19.

P200 Latency. Significant main effects were found for feedback condition, F (1, 11) = 7.40; and for group condition, F (2, 33) = 4.18. P200 latency to attended tones were longer when participants performed the auditory oddball task under the positive feedback condition (*M* = 220.91) than under the negative feedback condition (*M* = 213.19). Also, P200 latency was longer for participants in the adaptive automation group (*M* = 224.95) than for participants in the yoked condition (*M* = 212.95) or in the control condition (M = 213.25).

The results found for P200 latency for group condition must be viewed in consideration of the group X feedback interaction, F (2, 33) = 15.37. A simple effects analysis shows that only the adaptive automation, positive feedback combination (*M* = 239.19) was significantly different from the other group, feedback conditions. The other group, feedback condition combinations averaged approximately

Table 2. Means for ERP Components

Group	Feedback		
		N1 Amplitude	N1 Latency
a	p	-5.39	136.33
a	n	-3.60	140.16
y	p	-4.94	147.66
y	n	-3.35	142.00
c	p	-3.08	139.33
c	n	-3.21	141.91
		P2 Amplitude	P2 Latency
a	p	3.38	239.91
a	n	3.55	210.00
y	p	3.90	212.00
y	n	3.80	213.91
c	p	3.22	210.83
c	n	3.19	215.66
		P3 Amplitude	P3 Latency
a	p	1.75	350.41
a	n	4.40	306.91
y	p	1.99	348.75
y	n	2.20	331.00
c	p	2.10	338.00
c	n	2.18	329.66

Note. a = adaptive; y = yoke; c = control; n = negative; p = positive

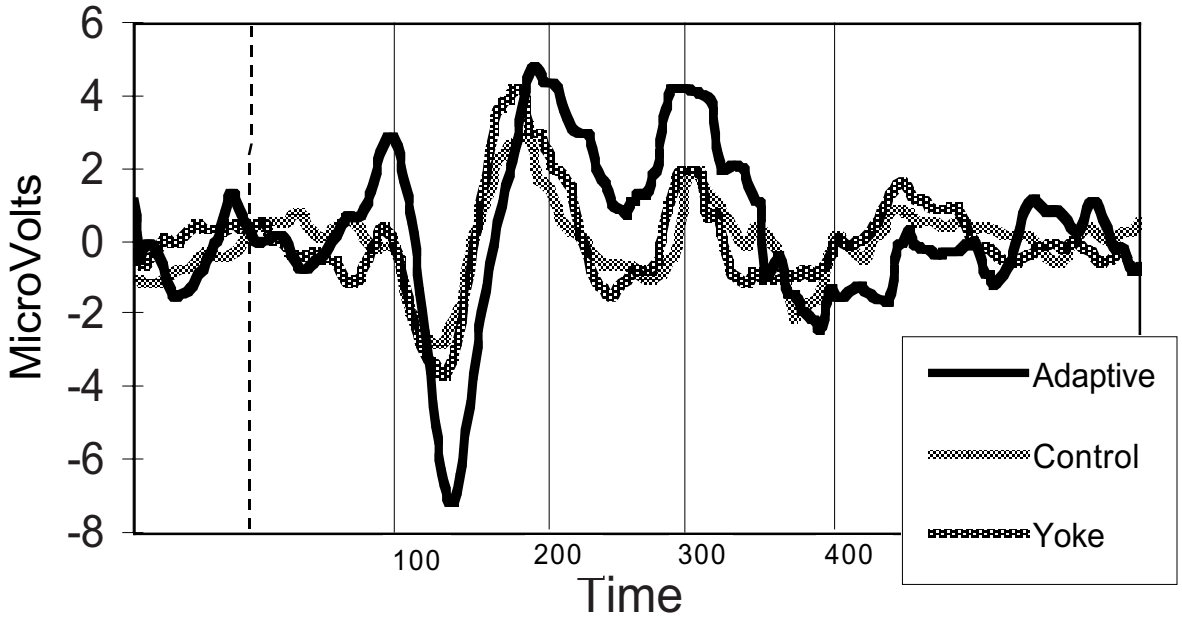


Figure 3. Overall ERP for Negative Feedback Contingency Across Groups

212 msec in latency. Therefore, the differences found for the main effect of group condition are due to the increased P200 latency in the positive feedback condition for participants in the adaptive automation group.

P300 Amplitude. An ANOVA yielded significant main effects for feedback condition, $F(1, 11) = 78.72$; and for group condition, $F(2, 33) = 20.40$. P300 amplitude was significantly larger when participants performed the task under the negative feedback condition ($M = 2.93$) than under the positive feedback condition ($M = 1.94$). Also, P300 amplitude was higher for those participants in the adaptive automation group ($M = 3.08$) than for those participants in the yoked condition ($M = 2.09$) or the control condition ($M = 2.14$). There was also a feedback condition X group interaction, $F(2, 33) = 57.21$. P300 amplitude was significantly higher under the negative feedback condition for participants in the adaptive automation group than under any other group, feedback combination.

P300 Latency. P300 latency was found to be significant only for feedback condition, $F(1, 33) = 13.91$. P300 latency was significantly longer under the positive feedback condition ($M = 345.72$) than under the negative feedback condition ($M = 322.52$). Neither group condition, $F(2, 33) = 0.99$; or group X feedback condition interaction, $F(2, 33) = 2.86$ were significant.

Discussion

Experiment Two was conducted to examine the efficacy of using event-related potentials and electroencephalogram for use in adaptive automation technology. Because psychophysiology is likely to be an essential aspect in the development of adaptive automation systems, it is necessary to research the issues that surround the use of these metrics. Furthermore, Experiment Two was design to add to the literature concerning the impact that adaptive automation has on behavioral, subjective, and ERP measures of workload and task engagement.

To accomplish these research goals, a multi-group design was used composed of adaptive automation, yoked, and control group conditions. Participants in the adaptive automation group were asked to perform a compensatory tracking task and an auditory oddball task while their EEG was continuously monitored. The tracking task was switched between manual and automatic task modes based upon whether their EEG was above or below baseline levels of task engagement and which feedback condition the system operated under. The automation schedule for each participant in the adaptive automation group was presented to a participant in the yoked condition. Therefore, each participant performed the tasks in the exact cycle sequence as their yoked counterpart. Additionally, a control group was employed that received a random assignment of task mode allocations.

The design was intended to enable the assessment of whether the adaptive automation method of task mode allocation represents a significantly better way of keeping operators "in-the-loop." If so, performance, subjective workload estimates, and psychophysiological correlates of workload would be better moderated for participants in the adaptive automation group, and no differences witnessed between the yoked or control group conditions. However, if adaptive automation does not significantly enhance the human-automation interaction, then no differences would be expected between the three experimental groups. Additionally, the design allowed for a determination to be made as to the utility of using EEG and ERPs in adaptive task allocation.

Experiment Two provided a wealth of data that has significant implications for adaptive automation design. Several significant results paralleled the findings from Experiment One. One of these results was for task allocations. As stated previously, if there were a functional relationship between the EEG

engagement index and task mode, the index should demonstrate stable short-cycle oscillation under negative feedback and longer and more variable periods of oscillation under positive feedback. The strength of the relationship would be reflected in the degree of contrast in the behavior of the index under the two feedback contingencies. This should be reflected in significantly more task allocations under the negative feedback condition than under the positive feedback condition. As predicted, there were more task allocations under negative feedback for both Experiment One and Two.

Another significant result found in Experiment Two that was similar to Experiment One was the performance and subjective workload results. Performance was significantly better and subjective workload was reported lower under negative feedback. Therefore, Experiment Two confirmed the conclusions from Experiment One that adaptive automation has the potential to improve operator performance and modulate mental workload. Because these results were discussed in Experiment One, no further discussion is provided. Instead, presented next is the significant results of interest of Experiment Two with regard to the efficacy of ERPs for adaptive automation design.

Event-Related Potentials

A number of researchers (Billings, 1997; Sheridan, 1997; Wickens, 1992; Wiener & Nagel, 1988) have noted that automation has changed the nature rather than reduced the workload demands placed on human operators. For example, pilots now focus on monitoring system controls and intervene only to detect, assess, and correct system failures. An important by-product of this role shift is the decreased ability to infer operator state because of limited interaction with the automated system. The use of advanced automation concepts, such as adaptive automation, would only increase such role transfer prompting the need for more diagnostic measures for the regulation of mental workload and other psychological constructs.

Byrne and Parasuraman (1996) discussed the role that various psychophysiological measures can play in the development of adaptive automation technology. They stated that ERPs possess a number of characteristics that make them ideal as candidate indices for adaptive task allocation. These include diagnostic specificity, sensitivity, and reliability (see Eggemeier, 1988). However, Parasuraman (Byrne & Parasuraman, 1996; Parasuraman, 1990) concluded that, although many proposals have been made concerning the use of ERPs in adaptive automation, little empirical evidence has been collected to support its efficacy.

The present study sought to address this limitation and assess whether ERPs can be used to make task allocations in an adaptive fashion. Specifically, it was designed to examine whether the ERP can discriminate between positive and negative feedback conditions. Furthermore, the study sought to determine whether differences were evident between the adaptive automation, yoked, and control group conditions in terms of ERP component waveforms. Finally, because any approach to adaptive automation requires multiple measures of operator state, another goal was to measure the degree of congruence that ERPs have with other workload metrics.

The ERP waveform components to the infrequent, high tones demonstrated significant differences in amplitude and latency between positive and negative feedback conditions. The P300 ERP component was significantly higher in amplitude under the negative feedback condition than under the positive feedback condition. Additionally, the P300 component was significantly shorter in latency under the negative feedback condition. These results support the findings for performance and subjective workload and demonstrate that the ERP was capable of discriminating between levels of task load in an adaptive environment. Therefore, they support other studies that have found that ERPs can be useful in the

development and application of adaptive automation technology (Kramer, 1991; Humphrey & Kramer, 1994; Trejo, Humphrey, & Kramer, 1996).

There was also an experimental group X feedback condition interaction for N100 and P300 amplitude. The adaptive automation, negative feedback condition produced P3s that were significantly larger in amplitude than any other group, feedback condition. The N100 was also found to be significantly higher in amplitude under the adaptive automation, negative feedback condition. There were no differences found between the yoked and control group conditions. Additionally, positive feedback for the adaptive automation group did not produce ERP waveforms that were significantly different from the yoked or control group conditions in either amplitude or latency measures.

Implications for Adaptive Automation

The results have implications for adaptive automation, particularly for mental models and resource allocation. In the following sections, the implications of both are presented.

Mental Models. The P300 is thought to index a context updating of our mental model of the environment (Donchin, Ritter, & McCallum, 1978). Donchin, McCarthy, Kutas, and Ritter (1983) stated that the P300 is a representation of neural action for updating the user's "mental model" that seems to underlie the ability of the nervous system to control behavior. The mental model then is an assessment of deviations from expected inputs and is, therefore, revised whenever discrepancies are found. The frequency of such revisions is dependent upon the "surprise value" and task relevance of the attended stimuli (e.g., high tones; Donchin, 1981). Therefore, the group X feedback condition interaction for P300 amplitude suggests that participants in the adaptive automation group may have been better able to predict the "state" of system operation, develop control strategies, select appropriate actions, and interpret the effects of selected actions (Gentner & Stevens, 1983; Johnson-Laird, 1983; Wickens, 1992; Wilson & Rutherford, 1989). The outcomes of such an improved mental model were improved performance and lowered workload and evidenced by larger amplitudes for the P300 ERP component.

Applications to Adaptive Automation. The recent interest in mental models is due to changing technology and there is a growing need for metaphors to describe the increasingly "black box" nature of systems (Howell, 1990; Wickens, 1992; Wilson & Rutherford, 1989). It is commonly accepted that people form mental models of tasks and systems, and that these models are used to guide behavior at the interface. Norman (1983) explains that people form internal, mental models of themselves and of the things with which they are interacting with. The extent to which the mental models provide a good fit determines whether users can understand the nature of this interaction. Therefore, automated processes must be made compatible with the users' internal representation of the system (Kantowitz & Campbell, 1996; Norman, 1983; Parasuraman & Riley, 1997; Scerbo, 1996).

The National Research Council (1982) further noted that the effectiveness of automation depends on matching the designs of automated systems to user's representations of the tasks they perform. The lack of a "match" between the operating characteristics of a system, the user's mental model of the system, and designer's conceptual model of the system can lead to increased errors, workload, response times, and so forth. As Kantowitz and Campbell (1996) suggest, automated design should provide timely, consistent, and accurate feedback, match task demands to environmental demands, design high stimulus-response compatibility, and develop appropriate operator training that facilitates the development of an accurate mental model.

The use of the mental model metaphor then is likely to be of continued service in the design of automated systems. Moreover, the development of advanced automation concepts should only increase the need for accessing the "black box" of the human operator. The need arises, therefore, for ways of measuring the degree of disparity between a user's mental model and the designer's conceptual model. The present results suggest that such can be supplied by the use of ERP measures although additional research would be needed to specify the nature of the ERP, its relation to user mental models, and how it could be used in adaptive automation design.

Resource Allocation. Another implication of these results concerns how the ERP relates to cognitive workload. As stated previously, the P300 is thought to represent the context updating of our mental model whenever a novel event occurs. Such an updating only occurs if the stimuli associated with a task requires that it be processed; that is, task-irrelevant stimuli that are ignored do not elicit a P300. However, consider the situation in which a participant is instructed to only partially ignore a stimulus, or a participant is asked to perform an oddball task while concurrently performing a tracking task as in the present study. Will the P300 measures reflect these graded changes in task difficulty? If so, then the P300 may serve as an index of the resource demands and, therefore, the cognitive workload imposed on the human operator (Gopher & Donchin, 1986; Kramer, 1987).

Research has consistently demonstrated that the P300 amplitude reflects the amount of expenditure of perceptual/central processing resources associated with performing a task(s) (Gopher & Donchin, 1986; Kramer, 1991; Parasuraman, 1990). The characteristics of the P300 exhibit a decrease in amplitude and an increase in latency to secondary task performance as the difficulty of the primary task is increased ("amplitude reciprocity hypothesis"; Isreal et al., 1977). The results of this study revealed that the P300 did indeed decrease in amplitude and increase in latency as the workload demands in the task increased. Furthermore, the group X feedback condition interaction for P300 supports the findings for performance and subjective workload and demonstrated that the use of adaptive task allocation reduced the workload for those participants performing the tasks in the negative feedback condition. In addition, the N100 and P200 waveforms further support the use of ERPs for adaptive automation because they are thought to represent the early processes of selective attention and resource allocation (Hackley, Woldoroff, & Hillyard, 1990; Hillyard, Hink, Schwent, & Picton, 1973).

Applications to Adaptive Automation. Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) argued that adaptive automation represents the coupling of levels of automation to levels of operator workload. Therefore, candidate indices, which serve as adaptive mechanisms, must be capable of discriminating between various levels of task load. Although a number of measures have been proposed, Morrison and Gluckman (1994) suggested the use of psychophysiological metrics because of their potential to yield real-time estimates of mental state with little or no impact on operator performance.

There are many CNS psychophysiological measures available to system designers seeking to use them in adaptive automation design, such as EEG, Transcranial Doppler, fMRI, PET, and functional near infrared tomography. However, because of the multidimensional nature of mental workload and other psychological constructs (e.g., memory, attention, language processes) that require attention in the design of automated systems, only the ERP to date has been found to be sensitive to these different information processing activities (Kramer, 1991; Kramer, Trejo, & Humphrey, 1996) although the efficacy of several other psychophysiological measures are being investigated. While the biocybernetic system did not predicate task allocation on the basis of ERP data, the results showed that the ERP was capable of discriminating between levels of taskload in an adaptive environment. Therefore, a next step would require the development of an adaptive algorithm that uses the components of the ERP waveform as an adaptive mechanism for allocating tasks between the operator and automated system. The research by

Humphrey and Kramer (1994) as well as the present results demonstrates that such a biopsychometric system is capable of development. Despite the fact that such a system may be years from fruition, at the very least these results demonstrate that the ERP can serve in the developmental role (see Byrne & Parasuraman, 1996) of adaptive automation design. Taken together, then, the results of the ERP data support the conclusion of many human factors professionals that ERPs possess the adaptive capabilities for determining optimal human-automation interaction (Byrne & Parasuraman, 1996; Defayolle et al., 1971; Donchin, 1980; Farwell & Donchin, 1988; Gomer, 1981; Kramer & Humphrey, 1994; Kramer, Humphrey, Sirevaag, & Mecklinger, 1989; Kramer, Trejo, & Humphrey, 1996; Sem-Jacobsen, 1981; Scerbo, 1996). However, additional research is needed to increase the sensitivity of the ERP components to information processing stages. Promising avenues are multi-measure approaches combining process imaging with ERPs to better discriminate across stages and levels of information processing and mental workload assessment. Therefore, research is being directed at other CNS and ANS measures of mental workload to determine their efficacy for real-time adaptive automation design.

Experiment Three

Various candidate psychophysiological measures are available for use in adaptive automation, depending on the application and the specific requirements for adaptation. For a recent review of psychophysiological measures for use in adaptive systems, see Scerbo et al. (2001). There are several advantages to such physiological measures (Byrne, & Parasuraman, 1996; Gomer, 1981; Parasuraman et al., 1992). In certain applications, these advantages may be sufficient to overcome the disadvantages of cost, user acceptance, etc. sometimes associated with the use of these measures. In Experiment Three, heart rate variability (HRV) was assessed for reasons of sensitivity, reliability, low cost, and ease of use (Fahrenberg & Wientjes, 2000), and to validate the biocybernetic diagnosticity for measures other than from central nervous system (CNS) etiology.

Experiment 3 was conducted to determine the efficacy of HRV for workload assessment and to determine HRV criteria on which to based real-time adaptive function allocation. In general, HRV decreases with increased workload demands. Because HRV had not been used previously in the adaptive system, a pre-experiment (Experiment 3a) was conducted to validate the diagnosticity of the measure for workload assessment before implementation in the biocybernetic system. Experiment 3b used results from the pre-experiment to develop the criteria for logic to determine adaptive function allocation decision in response to measured mental workload.

Experiment 3a used the EICAS-MAT to examine the sensitivity of heart rate measures to variations in task difficulty. The purpose of the study was to develop an appropriate, empirically-derived triggering algorithm based on variations in measured workload for use in an adaptive system in Experiment 3b. In order to develop relatively stable estimates of heart rate measures, three different task difficulty levels from low to high were used in a moderately large sample of young adults ($N = 30$).

Experiment 3a Method

Participants

Thirty young adults aged 18-25 participated, of whom 9 had some general aviation flight experience. All were right-handed, had 20/20 vision, and were not taking any medications affecting cardiovascular function. Participants were told to refrain from consumption of caffeinated drinks or foods for at least three hours prior to the study.

Task

The task set up was the version of the EICAS-MAT (see fig. 4) as used in the previous study by Parasuraman et al. (1999). Subjects performed three tasks simultaneously: (1) a two-dimensional compensatory tracking task requiring maintenance of a joystick controlled cursor over a central target area; (2) an engine systems monitoring task based on the EICAS display; and (3) a fuel management task requiring maintenance of fuel level in the aircraft supply tanks at specified levels. Participants performed all three tasks manually without any automation support in Experiment 3a. The EICAS task required an understanding of engine parameters such as the engine pressure ratio (EPR), exhaust gas temperature (EGT), etc. For the non-pilot participants, explanations of the EICAS variables and the typical abnormal values were provided prior to the task training and practice.

Figure 4. EICAS sub-window of Multi-Attribute Task Battery (printed in grayscale).

Heart Rate Recording

The electrocardiogram was recorded using a standard lead 1 configuration and was conditioned with an analog bandpass filter at 5-50 Hz. The conditioned signal was digitized at 1000 Hz for R-wave detection (to the nearest 1 ms). The resulting inter-beat intervals (IBI) were time sampled at 2 Hz. A moving polynomial filter was then applied to the time series to extract estimates of RSA and 0.1 Hz HRV based on variable length epochs (see Results section below). Additional details can be found elsewhere (Byrne, Chun, and Parasuraman, 1995; Masalonis, Duley, & Parasuraman, 1999).

Procedure

Three blocks of 30 minutes each were administered, following a 15-minute training session (approx. 30 min for non-pilots), a 10-minute practice session, and a resting baseline 5-minute session during which heart rate was recorded. Tracking difficulty was varied over three levels, low, medium, and high, by varying the bandwidth of the forcing function (0.05 Hz, .08 Hz, and 0.12 Hz, respectively). The order of tracking difficulty was counterbalanced across subjects. The NASA-TLX was administered following the initial practice block and after all three of the 30-minute blocks.

Results and Discussion

There were no significant differences in the performance patterns or heart rate values of participants with (N = 9) and without (N = 21) some flight experience. Hence the data are reported collapsed across all participants.

MAT Performance

For MAT performance, it was first ascertained whether variation in the forcing function bandwidth resulted in the expected changes in tracking performance and subjective workload. The analysis was conducted on the mean values of the tracking RMS error (in arbitrary pixel units) and the mean NASA-TLX scores (averaged over all sub-scales) for the low, medium, and high forcing function bandwidth conditions. Analysis of variance (ANOVA) showed that there was a highly significant effect of forcing function bandwidth on tracking RMS error ($p < .001$).

Tracking error more than tripled from the low (67.3) to the high (219.4) bandwidth condition. Post-hoc paired comparisons revealed significant differences across the three levels (medium = 147.1) of the independent variable of bandwidth. In addition, subjective perception of workload also increased significantly across the low (33.5), medium (45.5), and high (57.1) levels of tracking difficulty ($p < .01$).

Performance on the EICAS monitoring task and the fuel management task were also computed for each level of tracking difficulty. The detection rate on the EICAS task decreased slightly but significantly ($p < .05$) with tracking difficulty. Mean values for the low, medium, and high levels were 83.3%, 81.2%, and 78.4%. There were no significant effects of tracking difficulty on performance of the fuel management task, as assessed by averaged RMS error in fuel tank levels.

Heart Rate Measures

Having shown the validity of our tracking difficulty manipulation, we then examined the effects of variations in task demands on heart rate measures. First, baseline to task performance condition was compared. Next, the results compared heart rate measures (averaged over the 30 minute block) for the

different forcing function levels. Finally, as a prelude to Experiment 3b, the stability and reliability of heart rate measures was examined as a function of "window length", or the length of time over which the time series of IBI was evaluated. All heart rate measures were sensitive to task performance as compared to baseline. In particular, RSA and 0.10 Hz HRV were significantly ($p < .001$ in each case) lower during task performance than during baseline. Of greater interest was the change in heart rate parameters with task difficulty. The analysis was conducted on the mean values of RSA (in ln(ms)2) and 0.1 Hz HRV (also in ln(ms)2) as a function of forcing function bandwidth.

One-way ANOVAs were computed for the RSA and 0.1 Hz measures. For RSA, there was a trend towards reduction in RSA with forcing function bandwidth, but this tendency was not significant ($p < .15$). The mean values were low (6.11), medium (6.21), and high (5.89). For the 0.1 Hz measure, there was a significant effect of tracking difficulty ($p < .01$). Post-hoc tests of means showed that 0.1 Hz HRV was significantly lower in the high tracking (4.18) difficulty condition compared to the low (5.18; $p < .01$) and medium (5.27; $p < .05$) difficulty conditions, which did not differ significantly from each other. These findings indicate that the 0.1 Hz measure was sensitive both to the imposition of task demands (baseline to task performance) as well as to increases in tracking task difficulty. However, this heart rate index of workload was only sensitive when the highest level of tracking difficulty was included. Whereas the HRV measure differed between the high (bandwidth = .12 Hz) and the medium difficulty (bandwidth = .08 Hz) levels, it could not distinguish the medium and low difficulty levels. Accordingly, in Experiment 3b, tracking difficulty was varied between the low (bandwidth = .05 Hz) and high levels (bandwidth = .12 Hz).

Finally, the relative stability and reliability of the 0.1 Hz HRV measure as a function of "window length," or the time over which successive IBI samples were computed. Window length refers to the time history over which a measure of physiological state is assessed. Selection of an appropriate window length is important for both theoretical and practical reasons. It is intuitive that the window length should be neither too long nor too short. Theoretically, the window length should be sufficient to sample temporal changes in experienced workload in a prolonged, multi-task performance environment. A very long window length may not be sufficiently sensitive to momentary, large changes in workload. What about the other end of the temporal continuum? How short can the window be? A certain minimum window length is necessary for reliable extraction of the physiological parameter. However, beyond that minimum, too short a window length, if coupled with adaptive logic that implements changes in function allocation, could lead to unstable oscillations in system performance.

Accordingly, intra-subject reliability estimates of heart rate measures as a function of window length were computed. The basic procedure was to choose at random two 10-minute segments of time within a 30-minute block at one of the three levels of tracking difficulty. Given that only the 0.1 Hz component was sensitive to tracking difficulty changes, we analyzed only this heart rate measure. Within the first 10-minute segment, we computed values of 0.1 Hz HRV for several different window lengths, from 10 seconds to 200 seconds. It was then recomputed the measure for the next step at the same window length. Initially a step length of 10 seconds was chosen. The resulting time series (one for each window length) was then submitted to a non-linear curve fitting analysis to identify the point at which the measure stabilized. Figure 5 shows an example of the time series analysis for one participant for the 30-second window length. The 0.1 Hz HRV estimate was initially unstable, but reached asymptote to a stable value at a time length of between about 20 and 80 seconds.

The analysis was repeated for several step lengths, from 10 seconds to 100 seconds. Next, it was computed for the same estimates for the second 10-minute segment within the 30-minute block. We then

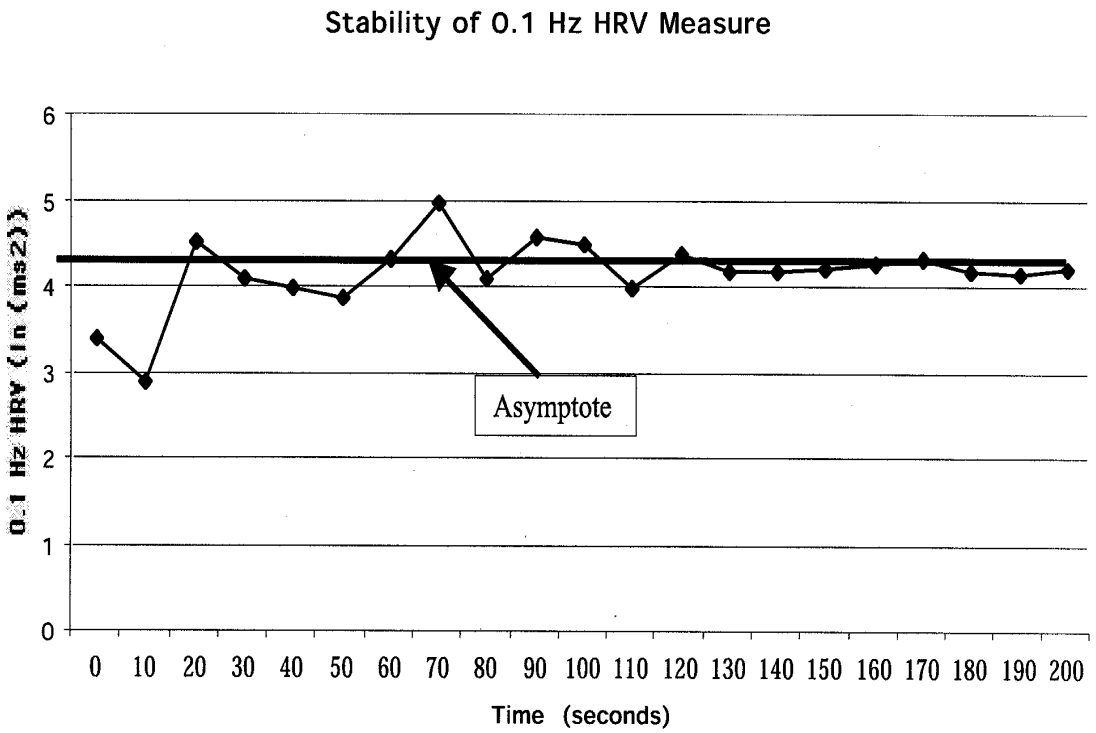


Figure 5. Within-Session Stability of 0.1 Hz HRV Estimate as a Function of Window Length.

computed test-retest correlations of these estimates across the two segments. Correlations were moderate to high, in the range of 0.45 - 0.85. Finally, a window length based on the stability analysis (as in fig. 5) was chosen and given a test-retest correlation of at least 0.8. This analysis led to the selection of a 30-second window length and a 30-second step length.

In summary, the results of Experiment 3a pointed to the feasibility of using heart rate measures to trigger adaptive automation, which was tested in Experiment 3b. Furthermore, the results validated the tracking difficulty manipulation, and showed that at least the extremes of the three levels tested were associated with reliable changes in 0.1 Hz HRV. Accordingly we chose these two levels (0.05 Hz and 0.12 Hz bandwidth) for the tracking difficult manipulation in Experiment 3b. Furthermore, the stability and reliability analysis showed that a 30-second window and a 30-second step length possessed features that would enhance the sensitivity of an HRV-based adaptive algorithm.

Experiment 3b utilized HRV estimates of workload, found in Experiment 3a, to develop real-time adaptive automation during performance of the MAT. The design of the study was similar to that of the previous study of Parasuraman et al. (1999). However, instead of adaptive changes being scripted to occur at specified times irrespective of individual subject workload, adaptation was keyed to measured workload in real time.

Experiment 3b Method

Participants

Twenty-six young adults aged 18-29 participated. All were right-handed, had 20/20 vision, and were not taking any medications affecting cardiovascular function.

Task

All participants first completed a 15-minute baseline condition in which they performed the MAT task while heart rate measures were computed. They then completed a 90-minute simulated flight session consisting of a three-phase (each 30 minutes), high-low-high task load profile typical of takeoff/climb, cruise, and approach/landing. The EICAS and the fuel management tasks were always performed manually and did not vary across flight phases or groups. Primary flight was carried out manually at all times, except when adaptive aiding (AA) was provided. AA consisted of automated lateral control of the tracking task. When AA was implemented (preceded and followed by clear, 30-second warning signals on the primary tracking display), subjects only had to track vertically.

Heart Rate-Based Adaptive Logic

Several parameters have to be defined for an effective adaptive algorithm or adaptive logic. These include the window length, the step length in a moving window system, pre-processing of input data to the adaptive logic (e.g, raw or normalized values), the number of values, and the decision logic, among others. As far as possible we attempted to choose parameters on an empirical basis, supplemented by practical limitations. Hence, it is not claimed that the particular adaptive logic chosen for implementation in Experiment 3b is necessarily the most effective one or optimal in any other sense. The "moving window" parameter was empirically determined. On the basis of the results of Experiment 3a and pre-experiment research, a value of 30 seconds (a value deemed optimal by previous research on psycho-physiological and adaptive automation) was used to assess the physiological state (heart rate measures) of individual subjects during each flight phase. The step length for the moving window (also determined empirically) was 30 seconds. The adaptive logic was as follows. During each 30-minute flight phase, the moving window estimate was computed and updated every 30 seconds. After a minimum of 5 minutes has elapsed into the flight phase, the estimates were assessed against the mean "transition point" values (for high and low workload) based on the estimates obtained in Study 1. Further details of the development of the adaptive algorithm are given in the Results section below. For example, if sometime during a particular phase of flight, the measured HRV was suppressed beyond the high workload transition point calculated for that subject, the adaptive logic would be triggered and AA (lateral hold automation) implemented in 30 seconds (following a warning message). If the transition point was not exceeded, the next estimate was computed after 30 seconds had elapsed, compared to the transition point value, the adaptive logic triggered or not, and so on. Once the adaptive logic was triggered, the moving window value was reset and a new moving window estimate was calculated and again updated every 30 seconds. The new value was now compared to the appropriate transition point value (e.g., low workload). This continued until 5 minutes remained in the flight phase. At that point the adaptive logic was disabled.

Accordingly, for each 30-minute flight phase, there was a period of 20 minutes when adaptive aiding could be triggered and/or control returned to the operator. Given that the moving window estimate of physiological state was updated every 30 seconds, and that a 30-second warning was given prior to a

change in automation status, the minimum time for the system to trigger and implement an adaptive change was 60 seconds. Hence, in principle, if the measured operator workload fluctuated and exceeded the transition points relatively rapidly, 20 adaptive changes in the automation status of the tracking task (lateral hold on or off) could be implemented in the 20 minutes of each flight phase when such changes were possible. One concern in the development of the adaptive logic was that such frequent changes in automation status might result in an unstable system and be perceived as disruptive to the establishment of a "work rhythm" by the participants. Following pilot work to develop the algorithm, however, and in practice, far fewer adaptive changes were triggered and implemented during each flight phase, as indicated in the Results section below.

Procedure

The 26 subjects were randomly assigned to one of two groups of 13 subjects each, an adaptive group and a nonadaptive control group. For the workload-matched adaptive group, AA was provided to the participants during the first and last 30-minute phases of the 90-minute session (high workload) whenever the heart rate based triggering algorithm exceeded the high-workload transition point. Conversely, AA was removed and full control of tracking returned to the participant when the low workload transition point was passed. For the control group, the task set up was the same, and heart rate was measured. However, no adaptive changes were triggered.

Results

Adaptive Transitions in Automation States

Despite our attempt to tailor the adaptive logic to the performance of individual participants by normalization and scaling of threshold values to individual baselines, the adaptive logic did not trigger completely in both flight phases for two of the 13 subjects in the adaptive group. For one of these participants, adaptive aiding was not triggered in either the first or the last flight phase of high-bandwidth tracking. For the other, AA was triggered in the first but not the last flight phase. Hence the data reported below are for the remaining 11 participants for whom AA was triggered at least once during each 30- minute flight phase. The mean performance levels of these two participants were in the top 20% of all participants. However, apart from the derived heart rate values showing reduced variation during the high-workload phase (and thus not triggering the adaptive logic), there was otherwise nothing remarkable about the performance or heart rate parameters of these two individuals.

For the remaining 11 participants in the adaptive group, the following analyses were conducted. At each point in time t during the first and last 30-minute flight phases, a moving window estimate of a parameter Ht, the derived 0.1 Hz HRV value, was obtained. This value was then evaluated against the transition point values associated with high and low workload. The mean 0.1 Hz HRV values for the high and low tracking difficulty levels from Experiment 3a were Thi = 4.18 and Tlo = 5.18 ln(ms)2 respectively. The individual participant's baseline 0.1 Hz HRV value was then used to scale these mean values to create individual-specific transition points for each participant. For example, the transition points for one participant with a baseline 0.1 Hz HRV value of 6.52 were Thi = 4.78 and Tlo = 5.92; whereas for another participant with a reduced baseline of 5.24 the corresponding points were Thi = 3.84 and Tlo = 4.76. Next, the current moving window estimate Ht was normalized with respect to the participant's baseline. Finally, the normalized estimate was thresholded against the transition point values, adjusted by a proportion k of the standard error (SE) of the current estimate. That is, if the current estimate of 0.1 Hz HRV was suppressed below the value Thi - kSE(Ht), adaptive aiding 20 was triggered (but not implemented for another 30 seconds). Conversely, once AA was implemented, if the current

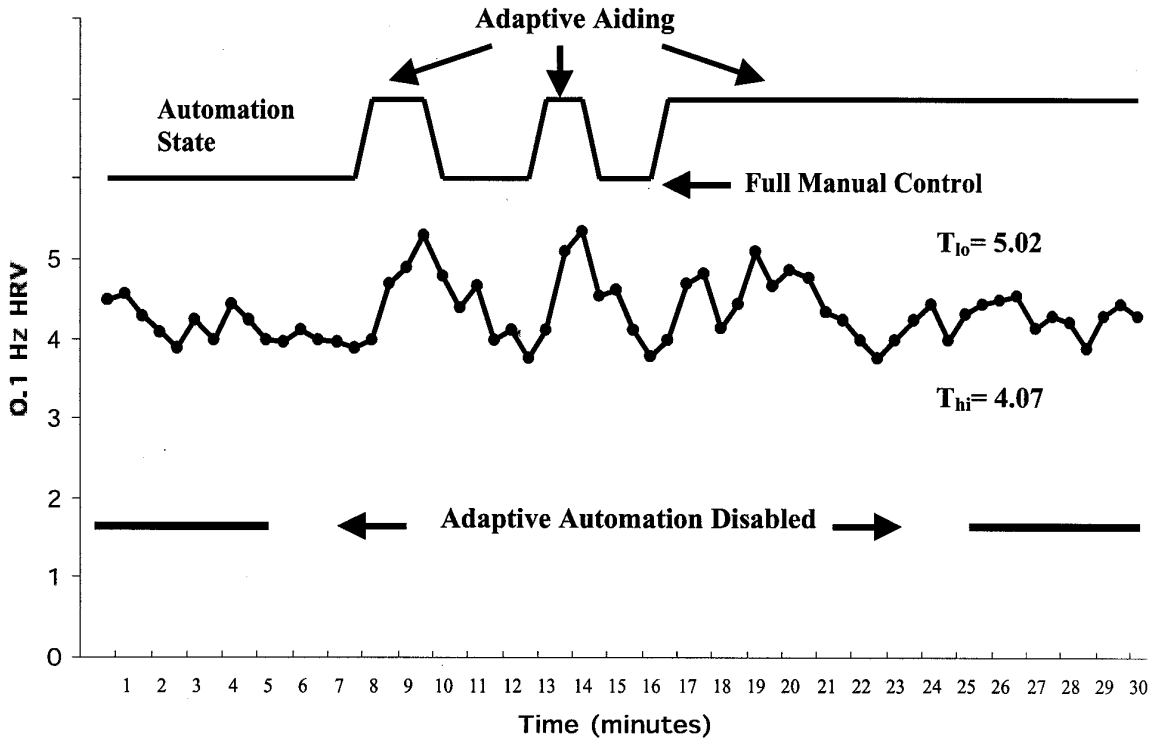


Figure 6. Mean 0.1 Hz HRV Values for an Individual Participant and Associated Transitions.

estimate was enhanced above the value *Tlo* + kSE(*H*t), control was returned to the participant. between Automation States. Transitions occurred at *Thi* - kSE(*H*t) and at *Tlo* + kSE(*H*t). Figure 6 shows the transition points for an individual participant in a single 30-minute flight phase. *Thi* and *Tlo* for this participant were 4.07 and 5.02, respectively. As figure 4 shows, this participant had three transitions to the automated state in which AA 1 For the initial analysis and implementation of AA in this study, *k* was set to 1. Additional analyses and future work will be necessary to determine an optimal value for *k.* (lateral hold automation) was implemented, and two transitions back to full manual control. The first transition (AA) took place at 8 minutes, or 3 minutes following the 5 minute initial period when the adaptive system was disabled, and following the triggering of the adaptive logic 7.5 minutes into the flight phase. Note that once only 5 minutes remained in the flight phase, no more transitions could take place, even if the HRV index was sufficient to trigger the adaptive logic.

Table 3 shows the mean numbers and durations of the adaptive changes, for both the high-low work-load transitions (when AA was implemented), and the low-high workload transitions (when the lateral hold was turned off and control returned fully to the participant). Data are shown separately for the first and last 30-minute flight phases when the tracking forcing function bandwidth was high. (Recall that adaptive changes were disabled for the first and last five minutes of each 30-minute phase, so these data are taken from the middle 20 minutes of each flight phase.) As Table 3 shows, the total number of transitions between automation states in the first phase ANOVA indicated that the mean number of adaptive function changes was significantly greater in the high-low then in the low-high direction ($p < .01$) and also declined from Phase 1 to Phase 3 ($p < .001$). There was no interaction between direction of change and phase. The mean duration of adaptive changes was also significantly greater for the high-low than for the low-high direction ($p < .001$) and also increased with flight phase ($p < .01$). In addition, there was a significant interaction between direction and phase ($p < .05$). As Table 3 shows,

Table 3. Mean Numbers and Durations of Adaptive Changes and Associated HRV During the First and Last 30-Minute Flight Phases

Direction of Transition	Phase 1		Phase 3	
	High - Low	Low - High	High - Low	Low - High
Number of Transitions Between Automation States	3.41	2.71	2.62	1.67
Duration of Transition State (minutes)	4.28	1.25	5.08	3.39
0.1 Hz HRV During Transition State	5.44	4.03	5.68	4.72

this was because the difference in transition durations between the two directions of adaptive function changes was reduced in Phase 3 compared to Phase 1. Thus, towards the end of the flight simulation, there were fewer but longer-duration transitions between automation states. Furthermore, the durations of the periods when AA was implemented (high-low) and withdrawn (low-high) came closer together in Phase 3 compared to Phase 1. This perhaps reflected the fact that by Phase 3 participants had greater experience of the adaptive system. More generally, the results validate the adaptive logic derived from the results of Study 1 and the pilot study and indicate not only that the resulting system performance did not oscillate too frequently, but that it also became relatively more stable with time.

Adaptive System Evaluation: Heart Rate and MAT Performance

The performance of the adaptive system in terms of heart rate measures and performance on the MAT task was also assessed. That the adaptive system performance was well linked to heart rate measures of workload is indicated by analysis of the mean 0.1 Hz HRV values during the transition states (see Table 3). The mean HRV value in the transition state for the high-low adaptive change (target state = low workload) was significantly greater ($p < .01$) than for the low-high adaptive change (target state = high workload). There was no significant effect of phase on these HRV values. The difference in HRV values between the two target automation states was somewhat lower in Phase 3 (mean difference in HRV = 1.41) than in Phase 1 (0.96). This trend could be indicative of the "leveling" effect of adaptive automation on workload and would be consistent with the greater stability of the adaptive system noted previously.

Unfortunately, the interaction between transition direction and flight phase was not significant ($p < .18$). Heart rate measures for the adaptive and the nonadaptive control groups were also compared across all three-flight phases. There were significant differences between the groups in mean 0.1 Hz HRV ($p < .05$) but not in RSA. HRV was higher in the adaptive group, indicating that workload was lower in this group compared to the control participants. Thus heart rate based adaptive automation not only reduced workload overall but also led to a greater leveling of workload between low and high tracking difficulty phases of flight.

The impact of the heart rate based adaptive system on multi-task performance was also examined, and the mean tracking RMS error for the adaptive and control groups was computed for all three phases of the flight simulation. As figure 7 shows, tracking performance was significantly ($p < .001$) higher in the adaptive group than in the control group. The main effect of phase ($p < .001$) and the interaction between phase and group ($p < .05$) were also significant. The main effect of group shows that the tracking performance of the adaptive group was consistently superior to that of the control group in all flight phases. The main effect of phase primarily reflects the reduction in tracking error in Phase 2 associated

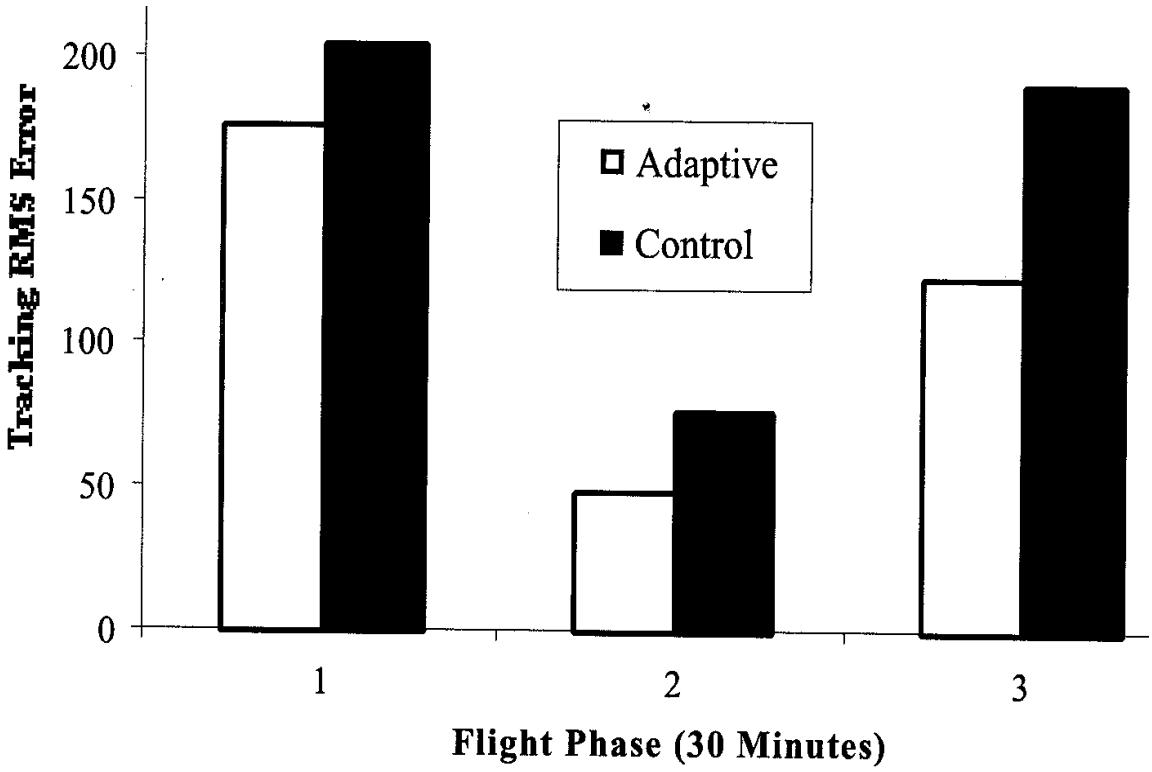


Figure 7. Tracking Performance as a Function of Flight Phase For the Adaptive and Control Groups.

with lower-bandwidth tracking. The interaction indicates that whereas the control group performance in the two high-bandwidth phases (1 and 3) did not differ, tracking error for the adaptive group was lower in Phase 3 than in Phase 1 (see fig. 7).

These results provide strong validation of the heart rate based approach to workload matching, and are consistent with the previous findings of Parasuraman et al. (1999). The results show that workload-matched adaptation is possible not only using a model-based approach, but also with real-time assessment of mental workload using heart rate variability measures. Importantly, the adaptive group was superior to the control group even in Phase 2, when no adaptive aiding was provided, suggesting a "carry-over" effect of prior adaptive automation in the previous flight phase. Parasuraman et al. (1996) reported similar persistent post-adaptive benefits on performance for the engine system-monitoring task of the MAT. Finally, that the adaptive group showed an improvement in performance following greater exposure to the adaptive system in the later flight phases is consistent with the “leveling” effect of adaptation on work-load noted in the earlier analysis of the number and duration of automation state transitions and the associated HRV values.

Performance on the EICAS monitoring task and the fuel management task were also computed for both groups and flight phases. The main effect of flight phase was significant for the detection rate on the EICAS task ($p < .01$). Detection rate was higher in Phase 2 (84.3%) than in Phase 1 (79.1%) or Phase 3 (77.3%). There were no significant differences in detection performance between groups. Finally, performance of the fuel management task did not differ between groups or phases.

Discussion

The results of these two studies validate a design approach to adaptive automation involving adaptation matched to operator mental workload (Hancock et al., 1985; Parasuraman et al., 1992). Adaptive aiding keyed to mental workload, as assessed by heart rate variability, led to an enhancement of performance in real time. In addition, heart rate based adaptive automation reduced overall workload and was also associated with a leveling workload between different phases of flight. The results provide strong support for the workload matching procedure proposed by Parasuraman et al. (1999). This approach to adaptive automation can be implemented using both a model-based approach as in that study and with physiological assessment of workload in real time as in the present study.

Since this was an initial study, no attempt was made to neither establish the efficacy of different adaptive algorithms nor establish in any sense the optimality of the chosen adaptive logic. Additional studies need to be conducted to examine other possible algorithms, including non-linear combination of parameters (e.g., with the use of a neural network model). Given that human-in-the-loop testing is time consuming and expensive, one avenue may be to conduct simulations of the effects of different possible adaptive algorithms. A model could be developed based on the performance and heart rate data obtained in the present. This model could then be used to simulate the impact of different algorithms with respect to parameters such as number and duration of transitions, stability, etc., as well as the potential impact on human performance.

Conclusions

Taken together, the results of these studies, coupled with other biocybernetic adaptive research (Hadley, Mikulka, Freeman, Scerbo, & Prinzel, 1997; Pope, Bogart, & Bartolome, 1995; Prinzel, et al., 1995; 1996; 1997; 1998; 2000; 2002) suggest that the closed-loop system represents a method for the use of psychophysiological measures in adaptive automation technology. However, because the closed-loop system has only been used for testing EEG, ERPs, and HRV indices, it remains to be seen whether other psychophysiological measures will also be appropriate for use with this system. Additionally, although these findings show potential for designing adaptive automation technology around psychophysiological measures, it may be some time before technology of this type becomes truly possible. Presently, psychophysiological recording technology offers few application possibilities outside of laboratory or clinical environments. Furthermore, the use of such measures suffers from a number of technical and theoretical shortcomings (Kramer, 1991; Byrne & Parasuraman, 1996). Also, general concerns about use, misuse, disuse, and abuse (Parasuraman & Riley, 1997) associated with the implementation of adaptive automation need to be considered. Nevertheless, adaptive automation represents one of the better ways of implementing automation (Mouloua & Parasuraman, 1994), and this form of technology offers one of the few direct applications of psychophysiology in the work environment (Byrne & Parasuraman, 1996). Presently, however, there is not enough existing psychophysiological research to provide adequate information on which to base adaptive allocation decisions. Although Byrne and Parasuraman suggested that some guidance can be found in other research domains, such as medical research (e.g., Martin, Schneider, Quinn, & Smith, 1992; Schwilden, Stoeckel, & Schuttler, 1989), more research is still needed that directly examines some of the special issues that surround the use of psychophysiological measures in adaptive automation.

The field of human factors has been traditionally defined as the design and evaluation of systems and tools for human use. The goal of human factors is directed at how people, machines, and the environment interact, and what can be done to make certain that productivity, efficiency, and safety are ensured. The idea that one should account for the human during the design process often seems too obvious to deserve

much attention. Recently, however, several known disasters and accidents have challenged such prevailing attitudes towards human factors research. The idea has certainly relevant for the use of automation especially in light of several disastrous accidents that have happened in the past few years in aviation transportation.

Scerbo (1996) noted that automation is neither inherently good nor bad. He stated that automation does, however, change the nature of work; it solves some problems while it creates others. Adaptive automation represents the next phase in the development of automated systems. To date, it is not known how this type of technology will impact work performance (Billings, 1997; Scerbo, 1996; Woods, 1996). However, it is clear that automation will continue to impact our lives requiring humans to co-evolve with the technology; this is what Hancock (1996) calls "techneology." Therefore, professionals involved with adaptive automation are incumbent to investigate the issues surrounding the use of adaptive automation technology. As Weiner and Curry (1980) conclude:

> The rapid pace of automation is outstripping one's ability to comprehend all the implications for crew performance. It is unrealistic to call for a halt to cockpit automation until the manifestations are completely understood. We do, however, call for those designing, analyzing, and installing automatic systems in the cockpit to do so carefully; to recognize the behavioral effects of automation; to avail themselves of present and future guidelines; and to be watchful for symptoms that might appear in training and operational settings (p.7)

The concerns they raised are as valid today as they were 18 years ago. Fortunately, at present, adaptive automation represents only a conceptual view of how automation can be advanced to improve the human-automation interaction. We now have an opportunity to research the technology before large-scale implementation of adaptive automation becomes available (Scerbo, 1996).

There are a number of issues that must be addressed before adaptive automation can move forward in the design of automated systems. To do otherwise, would be to risk repeating the fatal lessons of the past. As Billings and Woods (1994) noted,

> In high-risk, dynamic environments... technology-centered automation has tended to decrease human involvement in system tasks, and has thus impaired human situational awareness; both are unwanted consequences of today's system designs, but both are dangerous in high-risk systems. [At it's present state of development,] adaptive ("self-adapting") automation represents a potentially serious threat... to the authority that the human pilot must have to fulfill his or her responsibility for flight safety (p. 265).

Such a strong cautionary voice points to the need for more research in this area. The present study examined but a small share of these issues. These issues included the use of psychophysiological measures in adaptive automation design as well as a comparison of adaptive task allocation to static task allocation.

Byrne and Parasuraman (1996) stated that psychophysiology is an integral component of adaptive automation as a non-invasive method used to assess operator state. They suggested that such measures could be used not only as an input signal for the regulation of automation, but also to assess underlying changes accompanying performance changes during development of adaptive automation systems. The results support such a conclusion. The EEG, ERP and HRV were found to discriminate between positive and negative feedback controls and these were associated with other workload measures. Byrne and

Parasuraman noted that any psychophysiological measure must be used in conjunction with other metrics of operator state and any candidate indices must be capable of such an association. Indeed, these measures accorded well with the performance and subjective workload measures and, therefore, support Byrne and Parasuraman's assessment that biopsychometrics will play an important role in advanced automation.

Furthermore, these studies represent some of the few experiments, and the first on the use of ERPs and HRV, to demonstrate conclusively the advantages of the adaptive automation paradigm using a real-time approach. Parasuraman, Mouloua, & Molloy (1996) also examined the effects of adaptive task allocation, but they used model-based and performance-based approaches. These adaptive methods do not represent an adaptive aiding mechanism based on real-time measurements of operator workload. Furthermore, these researchers used only performance measures (i.e., reaction time, false alarms, hit rate, omissions). Kramer, Trejo, and Humphrey (1996) also examined the use of adaptive automation and provided both performance and psychophysiological measures. However, their study was a de facto assessment of how much ERP data is needed to discriminate different levels of mental workload and, therefore, was not adaptive automation in the truest sense. Therefore, the results reported here provide for the first controlled, empirical studies to evaluate the conjunctive effects of adaptive task allocation on behavioral, subjective, and psychophysiological correlates of workload.

Future Directions

Although the findings presented here give strong support for the benefits of adaptive automation and the use of psychophysiology in the design of this technology, the study only examined some of the many issues that need consideration. Parasuraman and his colleagues (Byrne & Parasuraman, 1996; Parasuraman, 1993; Parasuraman, Bahri, & Molloy, 1991; Parasuraman et al., 1992; Parasuraman, Mustapha, & Molloy, 1996) have noted a number of variables and factors that should be researched in adaptive automation design. These include the frequency of adaptive changes, adaptive algorithms, automation reliability and consistency, the type of interface, and contextual factors that are unique to specific systems. Scerbo (1996) also added system responsiveness, timing, and authority and invocation to this list. He further stated that research should branch out to other areas that are likely to be of concern for adaptive automation technology, such as mental models, teams, training, and communication. Moreover, if one considers the concerns of Woods (1996) that automation represents what he calls, "apparent simplicity, real complexity," one cannot leave without an impression that there is a considerable amount of work that is needed. However, research must begin somewhere and our work here, along with the works of others in the field are hoped to stimulate additional research in this new but exciting area of automation technology

References

Abbott, K., Slotte, S., Stimson, D., Amalberti, R. R., Bollin, G., Fabre, F., Hecht, S., Imrich, T., Lalley, R., Lydanne, G., Newman, T., & Thiel. G. (1996). *The interfaces between flightcrews and modern flight deck systems.* (Report of the FAA Human Factors Team). Washington DC: Federal Aviation Administration.

Alluisi, E. A., Coates, G. D., & Morgan, B. B., Jr. (1977). Effects of temporal stressors on vigilance and information processing. In R.R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates* (pp. 361-421). NY: Plenum.

Andreassi, J. L. (1995). *Psychophysiology: Human behavior & physiological response* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Backs, R. W., Lenneman, J. K., & Sicard, J. L. (1999). The use of autonomic components to improve cardiovascular assessment of mental workload in flight simulation. *The International Journal of Aviation Psychology, 9*, 33-47.

Backs, R. W., Ryan, A. M., & Wilson, G. F. (1994). Psychophysiological measures of workload during continuous manual performance. *Human Factors, 36*, 514-531.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, 775-779.

Barrett, G., Neshige, R., & Shibasaki, H. (1987). Human auditory and somatosensory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 66*, 409-419.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276-292.

Beatty, J. (1986). The pupillary system. In M.Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 43-50). New York: Guilford.

Beatty, J. (1988). Pupillometric signs of selective attention. In G. Galbraith, M. Kietzman, & E. Donchin (Eds.), *Neurophysiology and Psychophysiology* (pp. 85-110). Hillsdale, NJ: Lawrence Erlbaum, Associates.

Beatty, J., Greenberg, A., Diebler, W. P., & O'Hanlon, J. F. (1974). Operant control of occipital theta rhythm affects performance in a radar monitoring task. *Science, 183*, 871-873.

Beatty, J., & O'Hanlon, J. F. (1979). Operant control of posterior theta rhythms and vigilance performance: Repeated treatments and transfer of training. In N. Birbaumer and H. D. Kimmel (Eds.), *Progress in operant conditioning of physiological events* (pp. 235-246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Beck, E. C. (1975). Electrophysiology and behavior. *Annual Review of Psychology, 26*, 233-262.

Biferno, M. A. (1985). *Mental workload measurement: Event-related potentials and ratings of workload and fatigue.* (NASA CR-177354). Ames, CA: NASA.

Billings, C. E. (1991). *Human-centered aircraft automation philosophy: A concept and guidelines*. (NASA TM-103885). Moffett Field, CA: NASA.

Billings, C.E. (1997). *Automation aviation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 264-269). Hillsdale, NJ: Lawrence Erlbaum Associates.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety* (pp. 481-485). London: Taylor & Francis.

Byrne, E. A., Chun, K. M., & Parasuraman, R. (1995). Differential sensitivity of heart rate and heart rate variability as indices of mental workload in a multi-task environment. In *Proceedings of the 8th International Symposium on Aviation Psychology.* (pp. 881-885). Columbus, OH: Ohio State University.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*, 249-268.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of the mind: Event-related brain potentials and cognition* (pp. 1-26). Oxford: Oxford University Press.

Comstock, J. R., & Arnegard, R. J. (1991). *The multi-attribute task battery for human operator workload and strategic behavior research* (NASA Technical Memorandum No. 104174). Hampton, VA: NASA.

Davidson, R. J. (1988). EEG measures of cerebral asymmetry: Conceptual and methodological issues. *International Journal of Neuroscience, 39*, 71-89.

Davidson, F. J., Chapman, J. P., Chapman, L. J., & Henriques, J. B. (1990). Asymmetrical brain electrical activity discriminates between psychometrically matched verbal and spatial cognitive tasks. *Psychophysiology, 27*, 528-543.

Deecke, L. (1976). Potential changes associated with motor action, reflex responses, and readiness. In W. C. McCallum & J. R. Knott (Eds.), *The responsive brain* (pp. 91-93). Bristol, England: Wright.

Defayolle, M., Dinand, J., & Gentil, M. (1971). Averaged evoked potentials in relation to attitude, mental load, and intelligence. In W. T. Singleton, J. G. Fox, & D. Whitfield (Eds.), *Measurement of man at work* (pp. 141-152). London: Taylor & Francis.

De Weerd, J. P. C., & Kap, J. I. (1981). A posteriori time-varying filtering of averaged evoked potentials II: Mathematical and computational aspects. *Biological Cybernetics, 41*, 223-234.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked potentials and behavior* (pp. 13-75). New York: Plenum.

Donchin, E. (1980). Event-related potentials – Inferring cognitive activity in operational settings. In F.E. Gomer (Ed.), *Biocybernetic applications for military systems*. (McDonnell Douglas Technical Report MDC EB1911, pp. 35-42). Long Beach, CA: McDonnell Douglas.

Donchin, E. (1981). Suprise!...suprise? *Psychophysiology, 18*, 493-513.

Donchin, E., & Heffley, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potentials research* (pp. 555-572). Washington, D.C.: US Government Printing Office.

Donchin, E., Kramer, A. F., & Wickens, C. D. (1986). Applications of brain event-related potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 702-718). Middletown, NJ: Till & Till.

Donchin, E., McCarthy, G., Kutas, M., & Ritter, W. (1983). Event-related brain potentials in the study of consciousness. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Vol. 3* (pp. 81-121). New York: Plenum.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man* (pp. 247-265). New York: Academic Press.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 41-62). North-Holland: Elsevier Science Publishers.

Elbert, T., Rockstroh, B., Canavan, A., Birbaumer, N., Lutzenberger, W., von Bulow, I., & Linden, A. (1991). Self-regulation of slow cortical potentials an dits role in epileptogenesis. In J.G. Carlson & A.R. Seifert (Eds.), *International perspectives on self-regulation and health* (pp. 65-94). New York: Plenum.

Endsley, M. R. (1996). Automation and situational awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum, Assoc.

Endsley, M. R., & Kiris, E. O. (1994). The out-of-the-loop performance problem: Impact of level of automation and situation awareness. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 50-56). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fahrenberg, J., & Wientjes, C. J. E. (2000). Recording methods in applied environments. In R. W. Backs & W. Boucsein (Eds.), *Engineering Psychology: Issues and Applications* (pp. 111-136). London: Lawrence Erlbaum Associates.

Farwell, L., & Donchin, E. (1988). Talking off the top of your head: toward a mental prothesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology, 70*, 510-523.

Ford, J. M., Roth, W. T., & Koppell, B. S. (1976). Attention effects on auditory evoked potentials to infrequent events. *Biological Psychology, 4*, 65-77.

Fowler, B. (1994). P300 as a measure of workload during a simulated aircraft landing task. *Human Factors, 36*, 670-683.

Gaillard, A. W. K. (1993). Comparing the concepts of mental load and stress. *Ergonomics, 36*, 991-1005.

Gale, A. (1987). The electroencephalogram. In A. Gale and B. Christie (Eds.), *Psychophysiology and the electronic workplace* (pp.223-238). London: John Wiley and Sons.

Gale, A. & Christie, B. (1987). *Psychophysiology and the electronic workplace*. London: John Wiley and Sons.

Gale, A., & Edwards, J. (1983). The EEG and human behavior. In A. Gale & B. Christie (Eds.). *Physiological correlates of human behavior: Attention and performance* (pp. 99-127). New York: Academic Press.

Gentner, D., & Stevens, A. L. (Eds.) (1979). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Glenn, F., Barba, C., Wherry, R. J., Morrison, J., Hitchcock, E., & Gluckman, J. P. (1994). Adaptive automation effects on flight management task performance. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 33-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldstein, R., Walrath, L. C., Stern, J. A., & Strock, B. D. (1985). Blink activity in a discrimination task as a function of stimulus modality and schedule of presentation. *Psychophysiology, 22*, 629-635.

Gomer, F. (1981). Physiological monitoring ad the concept of adaptive systems. In J. Moraal & K.F. Kraiss (Eds.), *Manned systems design* (pp. 257-263). New York: Plenum Press.

Gopher, D., & Donchin, E. (1986). Workload – an examination of the concept. In K. Boff & L. Kaufman (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (pp. 41.1-41.49). New York: John Wiley.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Haagh, S. A. V. M., & Brunia, C. H. M. (1985). Anticipatory response-relevant muscle activity: CNV amplitude and simple reaction time. *Electroencephalography and Clinical Neurophysiology, 61*, 30-39.

Hackley, S., Woldoroff, M., & Hillyard, S. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem and cerebral evoked potentials. *Psychophysiology, 27*, 195-208.

Hadley, J., Mikulka, P. J., Freeman, F. G., Scerbo, M. W., & Prinzel III, L. J. (1997). *Psychophysiological evaluation of a biocybernetic system for adaptive automation*. Paper presented at the 9th Symposium on Aviation Psychology, Columbus, OH.

Hadley, J., Prinzel III, L. J., Freeman, F.G., & Mikulka, P. J. (in press 2003). Behavioral, subjective, and psychophysiological correlates of various schedules of short-cycle automation. In M. Scerbo & K. Krahl (Eds.), *Automation technology and human performance: Current research and trends*. Mahwah, NJ: Lawrence Erlbaum Associates.

Haider, M., Sprong, P., & Lindsley, D. B. (1964). Attention, vigilance, and cortical evoked potentials. *Science, 145*, 180-182.

Hancock, P. A. (1996). Teleology and technology. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (pp. 461-498). Mahwah, NJ: Lawrence Erlbaum Associates.

Hancock, P. A., & Chignell, M. H. (1987). Adaptive control in human-machine systems. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 305-345). North Holland: Elsevier Science Publishers.

Hancock, P. A., Chignell, M. H., & Lowenthal, A. (1985). An adaptive human-machine system. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 15*, 627-629.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, 18*, 647-658.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Hilburn, B., Jorna, P. G. A. M., & Parasuraman, R. (1995). The effects of advanced ATC automation on mental workload and monitoring performance: An empirical investigation in Dutch airspace. *Proceedings of the 8th international symposium on aviation psychology* (pp. 387-391). Columbus OH, Ohio State University.

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). *Operator versus computer control of adaptive automation* (Technical Report NAWACADWAR-93031-60). Naval Air Warfare Center, Warminster, PA.

Hilburn, B., Parasuraman, R., & Mouloua, M. (1995). Effects of short- and long-cycle adaptive function allocation on performance of flight-related tasks. In N. Johnston, R. Fuller, & N. McDonald (Eds.), *Aviation psychology: Training and selection* (pp. 347-353). Hampshire, England: Ashgate.

Hillyard, S. A., & Picton, T. W. (1979). Event-related brain potentials and selective information in man. In J.E. Desmedt (Ed.), *Cognitive components in cerebral event-related potentials and selective attention – Progress in clinical neurophysiology* (pp. 110-187). Basel: Karger.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*, 177-180.

Hockey, G. R. L., Coles, M. G. H., & Gaillard, A. W. K. (1986). Energetic issues in research on human information processing. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 3-21). Dordrecht: Kluwer.

Hoffman, J. E. (1990). Event-related potentials and automatic and controlled processes. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 145-157). New York: Oxford University Press.

Hoffman, J. E., Houck, M., MacMillan, F., Simons, R., & Oatman, L. (1985). Event-related potentials elicited by automatic targets: A dual-task analysis. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 50-61.

Hoffman, J. E., Simons, R. F., & Houck, M. R. (1983). The role of attentional resources in automatic detection. *Cognitive Psychology, 15*, 625-631.

Horst, R. L. (1987). An overview of current approaches and future challenges in physiological monitoring. *Mental State Examination* (NASA CP-2504, pp.25-42).

Horst, R. L., Johnson, R., & Donchin, E. (1980). Event-related brain potentials and subjective probability in a learning task. *Memory & Cognition, 8*, 476-488.

Howell, W.C. (1990). Human factors in the workplace. In M.Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 209-269). Palo Alto, CA: Consulting Psychologist Press, Inc.

Humphrey, D. G., & Kramer, A. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. *Human Factors, 36*, 3-26.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980a). P300 and tracking difficulty: Evidence for multiple resources in dual task performance. *Psychophysiology, 17*, 57-70.

Isreal, J. B., Wickens, C. D., Chesney, G.L., & Donchin, E. (1980b). The event-related brain potential as an index of display-monitoring workload. *Human Factors, 22*, 211-224.

Itoh, Y., Hayashi, Y., Tsukui, I., & Saito, S., (1989). Heart rate variability and subjective mental workload in flight task validity of mental workload measurement using HRV method. In M. J. Smith, & G. Salvendy (Eds.), *Work with computers: Organizational, management, stress, and health aspects* (pp. 209-216). Amsterdam: Elsevier Science Publishers.

Janssen, W. H., & Gaillard, A. W. K. (1985). EEG and heart rate correlates of task load in car driving. In A. Gundel (Ed.), *Proceedings of the workshop: Electroencephalography in transport operations* (pp. 39-53). Cologne, Germany: Elsevier.

Jasper, H. (1958). Report on the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology, 10*, 370-375.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*, 367-384.

Johnson, R., Miltner, W., & Braune, C. (1991). Auditory and somatosensory event-related potentials: I. Effects of attention. *Journal of Psychophysiology, 5*, 11-25.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. *Biological Psychology, 34*, 237-257.

Kaber, D. B., & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on workload assessment through a secondary monitoring task. In M. Scerbo & M. Mouloua (Eds.) *Automation technology and human performance: Current research and trends*. (pp. 129-133). Mahwah, NJ: Erlbaum.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Peavler, W.S. (1969). Insensitive effects and pupillarity changes in association learning. *Journal of Experimental Psychology, 79*, 312-318.

Kahneman, D., Peavler, W. S., & Onuska, L. (1968). Effects of verbalization on incentive on the pupil response to mental activity. *Canadian Journal of Psychology, 22*, 186-196.

Kahneman, D., Tversky, B., Shapiro, D., & Crider, A. (1967). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 27*, 187-196.

Kantowitz, B. H., & Campbell, J. L. (1996). Pilot workload and flightdeck automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Kramer, A. F. (1985). The interpretation of the component structure of event-related brain potentials. *Psychophysiology, 22*, 334-344.

Kramer, A. F. (1987). Event-related brain potentials. In A. Gale, & B. Christie (Eds.), *Psychophysiology and the electronic workplace* (pp. 197-222). New York: John Wiley & Sons.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-Task Performance* (pp. 279-328). London: Taylor & Francis.

Kramer, A. F., Humphrey, D., Sirevaag, E., & Mecklinger, A. (1989). Real-time measurement of mental workload: A feasibility study. *Proceedings of the third annual workshop on space operations, automation, and robotics* (pp. 83-102). Houston, TX: NASA Johnson Space Center.

Kramer, A. F., Schneider, W., Fisk, A., & Donchin, E. (1986). The effects of practice and task structure on components of the event-related brain potential. *Psychophysiology, 23*, 33-47.

Kramer, A. F., Sirevaag, E., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors, 29*, 145-160.

Kramer, A. F., & Strayer, D. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology, 26*, 231-268.

Kramer, A. F., Trejo, L. J., & Humphrey, D. G. (1996). Psychophysiological measures of workload: Potential applications to adaptively automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 137-162). Mawhaw, NJ: Lawrence Erlbaum Associates.

Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. In J. T. Cacioppo, L. G., Tassinary, & Berntson, G. G. (Eds.) *Handbook of psychophysiology. 2nd edition.* (pp. 794-814). New York: Cambridge University Press.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing requirements of a complex perceptual-motor task. *Human Factors, 25*, 597-621.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 393-408.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*, 792-795.

Lindholm, E., & Cheatham, C. M. (1983). Autonomic activity and workload during learning of a simulated aircraft carrier landing task. *Aviation, Space, and Environmental Medicine, 54*, 435-439.

Lindholm, E., & Cheatham, & C. M., Koriath. (1984). *Physiological assessment of aircraft pilot workload in simulated landing and simulated hostile threat environments*. (AFHRL-TR-83-49). Williams Air Force Base, Arizona: US Air Force.

Lindqvist, A., Keskinen, E., Antila, K., Halkola, L., Peltonen, T., & Valimaki, I. (1983). Heart rate variability, cardiac mechanics, and subjectively evaluated stress during simulator flight. *Aviation, Space, and Environmental Medicine, 54*, 685-690.

Lubar, J. F. (1991). Discourse on the development of EEG diagnostics and biofeedback for Attention-Deficit/ Hyperactivity Disorders. *Biofeedback and Self-Regulation, 16*, 201-225.

Masalonis, A. J., Duley, J. A., & Parasuraman, R. (1999). Effects of manual and autopilot control on mental workload and vigilance during general aviation imulated flight. *Transportation Human Factors, 1,* 187-200.

McCallum, C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 59*, 477-488.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*, 77-79.

McGillem, C. D., & Aunon, J. I. (1977). Measurements of signal components in single visually evoked brain potentials. *IEEE Transactions in Biomedical Engineering, BME-24*, 232-241.

Morris, N. M., & Rouse, W. B. (1986). *Adaptive aiding for human-computer control: Experimental studies of dynamic task allocation* (Report AAMRL-TR-86-005). Dayton, OH: Wright-Patterson Air Force Base.

Morrison, J. G., & Gluckman, J. P. (1994). Definitions and prospective guidelines for the application of adaptive automation. In M.Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 256-263). Hillsdale, NJ: Lawrence Erlbaum, Assoc.

Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1991). *Program plan for the adaptive function allocation for intelligent cockpits (AFAIC) program*. (Final Report No. NADC-91028-60). Warminster, PA: Naval Air Development Center.

Mouloua, M., & Parasuraman, R. (Eds.) (1995). *Human performance in automated systems: Current research and trends*. New Jersey: Lawrence Erlbaum, Assoc.

Naatanen, R., & Picton, T. W. (1987). The N1 wave fo the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*, 375-425.

Natani, K., & Gomer, F. E. (1981). Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and in event-related potentials. *(McDonnell Douglas Technical Report E2427)*. Long Beach, CA: McDonnell Douglas Corporation.

National Research Council (1982). *Automation in combat aircraft*. Washington, D.C.: National Academy Press.

Neville, H. J. (1985). Biological constraints on semantic processing: A comparison of spoken and signed languages. *Psychophysiology, 22*, 576.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Donnell, R., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. II: Cognitive processes and performance* (pp. 42-1 – 42-47). New York: John Wiley.

Offenloch, K., & Zahner, G. (1990). Computer aided physiological assessment of the functional state of pilots during simulated flight. *NATO Advisory Group for Aerospace Research and Development Conference Proceedings, 490*, 9-1 to 9-9.

O'Hanlon, J. F., & Beatty, J. (1979). Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance. In R. R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates* (pp. 189-201). NY: Plenum.

O'Hanlon, J. F., & Beatty, J. (1977). Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance. In R. R. Mackie (Ed.), *Vigilance* (pp. 189-201). New York: Plenum Press.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97-113.

Opmeer, C. H. J. M., & Krol, J. P. (1973). Towards an objective assessment of cockpit workload: I. Physiological variables during different flight phases. *Aerospace Medicine, 44*, 527-532.

Parasuraman, R. (1990). Event-related brain potentials and human factors research. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 279-300). New York: Oxford University Press.

Parasurman, R. (1993). Effects of adaptive function allocation on human performance. In D. J. Garland & J. A. Wise (Eds.), *Human factors and advanced aviation technologies* (pp. 147-157). Dayton Beach, FL: Embry-Riddle Aeronautical University Press.

Parasuraman, R. (1993). Adaptive function allocation and human performance. In D. Garland (Ed.) *Proceedings of the FAA/NASA Conference on Artificial intelligence and Human Factors in Air-Traffic Control and Aviation maintenance*. Daytona Beach, FL: Embry-Riddle Aeronautical University.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics, 43*, 931-951.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in adaptive systems* (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R., & Byrne, E. A. (2002). Automation and human performance in aviation. In P. Tsang & M. Vidulich (Eds.) *Principles and practice of aviation psychology.* Mahwah, NJ: Erlbaum.

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of workload. In P. A. Hancock and P. Desmond (Eds.), *Stress, workload, and fatigue.* (pp. 305-320). Mahwah, NJ: Erlbaum.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency." *International Journal of Aviation Psychology, 3*, 1-23.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. Scerbo & M. Mouloua (Eds.) *Automation technology and human performance: Current research and trends*. (pp. 119-123). Mahwah, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*, 665-679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Parsons, H. M. (1985). Automation and the individual: Comprehensive and comparative views. *Human Factors, 27*, 99-112.

Picton, T. W. (1988). Introduction. In T.W. Picton (Ed.), *Handbook of encephalography and neurophysiology: Human event-related potentials* (pp. 1-5). Amsterdam: Elsevier.

Pope, A. T., Bogart, E. H., & Bartolome, D. S (1995). Biocybernetic system validates index of operator engagement in automated task. *Biological Psychology, 40*, 187-195.

Prinzel III, L. J., Hitt, J., Scerbo, M. W., & Freeman, F. G. (abstract, 1995). Feedback contingencies and biocybernetic regulation of operator workload. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 39*, 947.

Prinzel III, L. J., Scerbo, M. W., Freeman, F. G, & Mikulka, P. J. (1995). A biocybernetic system for adaptive automation. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 39*, 1365-1369.

Prinzel III, L. J., Freeman, F. G., Scerbo, M. W., & Mikulka, P. J. (1997). A biocybernetic system for examining psychophysiological correlates for adaptive automation. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 41*, 1378.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive automation. *International Journal of Aviation Psychology, 10,* 393-410.

Prinzel III, L. J., Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (1997). Behavioral and physiological correlates of a bio-cybernetic, closed-loop system for adaptive automation. In M. Mouloua & J. M. Koonce (Eds.), *Human-Automation interaction: Research and practice* (pp. 66-75). Mahwah, NJ: Lawrence Erlbaum Associates.

Qiyuan, J., Richer, F., Wagoner, B., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology, 22*, 530-534.

Ragot, R. (1984). Perceptual and motor space representation: An event related potential study. *Psychophysiology, 21*, 159-170.

Ray, W., & Cole, H. (1985). EEG alpha activity during cognitive processing: Influence of attentional factors. *International Journal of Psychophysiology, 3*, 43-48.

Rebert, C. S., & Tecce, J. J. (1973). A summary of CNV and reactio time. IN W.C. McCallum & J.R. Knott (Eds.), *Event-related slow potentials of the brain: Their relations to behavior* (pp. 173-178). Amsterdam: Elsevier.

Redondo, M., & Del Valle-Inclan, F. (1992). Decrements in heart rate variability during memory search. *International Journal of Psychophysiology, 18*, 477-482.

Reising, J. M. (1985). 2010: The symbionic cockpit. *Proceedings of the IEEE National Aerospace and Electronics Conference, 3*, 1050-1054.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology, 22*, 204-207.

Richer, F., & Beatty, J. (1987). Contrasting effects of response uncertainty on the task evoked pupillary response and reaction time. *Psychophysiology, 24*, 258-262.

Richer, F., Silverman, C., & Beatty, J. (1983). Response selection and initiation in speeded reactions: A pupillometric analysis. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 360-370.

Riley, V. (1985). Monitoring the monitor: Some possible effects of embedding human models in highly automated manned systems. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 15*, 6-9.

Riley, V. (1994). A theory of operator reliance of automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 8-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Riley, V., Lyall, B., & Wiener, E. (1993). *Analytic methods for flight-deck automation design and evaluation. Phase two report: Pilot use of automation*. Minneapolis, MN: Honeywell Technology Center.

Rouse, W. B. (1977). Human-computer interaction in multitask situations. *IEEE Transactions Systems, Man, and Cybernetics, SMC-7*, 293-300.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30*, 431-443.

Rouse, W. B. (1991). *Design for success: A human centered approach to designing successful products and systems*. New York: John Wiley.

Rouse, W. B. (1994). Twenty years of adaptive aiding: Origin of the concept and lessons learned. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 28-32). Hillsdale, NJ: Lawrence Erlbaum, Assoc.

Ruchkin, D. S. (1988). Measurement of ERPs. In T. W. Picton (Ed.), *Handbook of encephalography and neurophysiology: Human event-related potentials* (pp. 7-43). Amsterdam: Elsevier.

Ruchkin, D. S., Sutton, S., & Teuting, P. (1975). Emitted and evoked P300 potentials and variation in stimulus probability. *Psychophysiology, 12*, 591-595.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22*, 642-647.

Sanders, M. S. & McCormick, E. J. (1993). *Human Factors in engineering and design* (7th Edition). NY: McGraw-Hill.

Sarter, N., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology, 4*, 1-28.

Sarter, N., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.) *Handbook of human factors and ergonomics*. 2nd ed. (pp. 1926-1943). New York: Wiley.

Sayers, B. M. C. A. (1973). Analysis of heart rate variability. *Ergonomics, 16*, 17-32.

Scallen, S. F., Hancock, P. A., & Duley, J. A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics, 26*, 397-403.

Scallen, S. F., Duley, J. A., & Hancock, P. A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 154-160). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scerbo, M. W. (1994). Implementing adaptive automation in aviation: The pilot-cockpit team. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 249-255). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 37-64). Mahwah, New Jersey: Lawrence Erlbaum Assoc.

Scerbo, M. W. (2001). Adaptive automation. In W. Karwowski (Ed.) *International encyclopedia of ergonomics and human factors*. London: Taylor and Francis, Inc.

Scerbo, M., Freeman, F., Mikulka, P. J., Di Nocera, F., Parasuraman, R., & Prinzel, L. (2001). *The efficacy of physiological measures for implementing adaptive technology.* NASA TP-2001-211018. Hampton, VA: NASA Langley Research Center.

Scherg, M., Vajsar, J., Picton, T. W. (1989). A source analysis of the late human auditory evoked potential. *Journal of Cognitive Neuroscience, 1*, 336-355.

Schwilden, H., Stoeckel, H., & Schuttler, J. (1989). Closed-loop feedback control of propofol anaesthesia by quantitative EEG analysis in humans. *British Journal of Anesthesiology, 62*, 290-296.

Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, Y., & Kuroda, I. (1978). Evaluation method of mental workload under flight conditions. *Aviation, Space, and Environmental Medicine, 49*, 920-925.

Sem-Jacobsen, C. W. (1981). Brain/computer communication to reduce human error: A perspective. *Aviation, Space, and Environmental Medicine, 52*, 31-38.

Sheridan, T. B. (1997). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1295-1327). New York: John Wiley & Sons.

Sherg, M. (1990). Fundamentals of dipole soure potential analysis. In F. Grandori, M. Hoke, & G.L. Romani (Eds.), *Auditory evoked magnetic fields and electric potentials. Advances in audiology* (pp. 40-69). Karger, Basel.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *International Journal of Aviation Psychology, 3*, 111-121.

Sirevaag, E., Kramer, A., Coles, M. G. H., & Donchin, E. (1989). Resource reciprocity: An event-related brain potential analysis. *Acta Psychologia, 70*, 77-97.

Sirevaag, E., Kramer, A., deJong, R., & Mecklinger, A. (1988). A psychophysiological analysis of multi-task processing demands. *Psychophysiology, 25*, 482.

Sirevaag, E., Kramer, A., Wickens, C. D., Reisweber, M., Strayer, D., & Grenell, J. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics, 36*, 1121-1140.

Smith, T. J., & Smith, K. U. (1987). Feedback control mechanisms of human behavior. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 251-293). New York: John Wiley & Sons.

Spyker, D., Stackhouse, A., Khalafall, A., & McLane, R. (1971). *Development of techniques for measuring pilot workload*. (NASA CR-1888). Minneapolis, MN: Honeywell Technology Center.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*, 387-401.

Sterman, M. B., Kaiser, D. A., Mann, C. A., Suyenobu, B. Y., Beyma, D. C., & Francis, J. R. (1993). Application of quantitative EEG analysis to workload assessment in an advanced aircraft simulator. *Proceedings of Annual Meeting of the Human Factors and Ergonomics Society, 37*,118-121.

Sterman, M. B., & Mann, C. A. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological Psychology, 40*, 115-130.

Sterman, M. B., Schummer, G. J., Dushenko, T. W., & Smith, J. C. (1987). *Electroencephalographic correlates of pilot performance: Simulation and in-flight studies* (pp. 1-16). Neuilly sur Seine, France: AGARD.

Straube, E., Schlenker, R., Klessinger, T., Himer, W., & Boven, K. (1987). Electrodermal orienting response and selective attention. *Psychophysiology, 24*, 615.

Strayer, D., & Kramer, A. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 67-82.

Streitberg, B., Röhmel, J., Herrmann, W. M., & Kubicki, S. (1987). COMSTAT rule for vigilance classification based on spontaneous EEG activity. *Neuropsychobiology, 17*, 105-117.

Surwillo, W. W. (1990). *Psychophysiology for clinical psychologists.* Norwood, NJ: Ablex publishing company.

Sutton, S., Braren, M., Zubin, J., John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science, 150*, 1187-1188.

Sutton, S., Teuting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science, 155*, 1436-1439.

Tattersall, A. J., & Hockey, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors, 37,* 682-698.

Teece, J. J. (1972). Contingent negative variation. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (pp. 657-679). Baltimore: Urban & Schwarzenberg.

Teece, J. J., & Cattanach, L. (1993). Contingent negative variation. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (pp. 887-910). Baltimore: Urban & Schwarzenberg.

Teuting, P., Sutton, S., & Zubin, J. (1970). Quantitative evoked potential correlates of the probability of events. *Psychophysiology, 7*, 385-394.

Theissen, M. S., Lay, J. E., & Stern, J. A. (1986). *Neuropsychological workload test battery validation study*. (FZM 7446). Fort Worth, TX: General Dynamics.

Torsvall, L., & Akerstedt, T. (1987). Sleepiness on the job: Continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology, 66*, 502-511.

Ullsperger, P., Metz, A., & Gille, H. (1988). The P300 component of the event-related brain potential and mental effort. *Ergonomics, 31*, 1127-1137.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41,* 656-669.

Vaughan, H. G., & Arezzo, J. C. (1988). The neural basis of event-related potentials. In T.W. Picton (Ed.), *Handbook of encephalography and neurophysiology: Human event-related potentials* (pp. 45-96). Amsterdam: Elsevier.

Vossel, G., & Rossmann, R. (1984). Electrodermal habituation speed and visual monitoring performance. *Psychophysiology, 21*, 97-100.

Wesensten, N. J., Crowley, J., Balkin, T., & Kamimori, G. (1993). Effects of simulated high altitude exposure on long-latency event-related brain potentials and performance. *Aviation, Space, and Environmental Medicine, 64*, 30-36.

Wickens, C. D. (1990). Applications of event-related potential research to problems in human factors. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 301-309). New York: Oxford University Press.

Wickens, C. D. (1992). *Engineering psychology and human performance, 2nd Ed*. New York: Harper Collins.

Wickens, C. D., Isreal, J. B., & Donchin, E. (1977). The event related cortical potential as an index of workload. *Proceedings of the Human Factors and Ergonomics Society, 21*, 282-287.

Wickens, C. D., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychological analysis of the reciprocity of information processing resources. *Science, 221*, 1080-1082.

Wickens, C. D., Mavor, A., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control: Human operators and automation.* Washington DC: National Academy Press.

Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D.C. Nagel (Eds.), *Human factors in aviation* (pp. 433-461). San Diego: Academic Press.

Wiener, E. L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (NASA CR-117528). Moffett Field, CA: NASA-Ames Research Center.

Wiener, E. L., & Curry, R. E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23*, 995-1011.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*. New York: J. Wiley & Sons.

Wilson, G. F. (1987). Electrodermal lability and simple reaction time. *Biological Psychology, 21*, 97-100.

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology, 34*, 163-178.

Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-Task Performance* (pp. 329-360). London: Taylor & Francis.

Wilson, G. F., & Graham, S. (1989). Electrodermal lability and visual information processing. *Psychophysiology, 26*, 321-328.

Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and applications in human factors. *Human Factors, 31*, 617-633.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Assoc.

Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical Biological Engineering, 5*, 529-553.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 09 - 2003	Technical Publication	

4. TITLE AND SUBTITLE

Three Experiments Examining the Use of Electroencephalogram, Event-Related Potentials, and Heart-Rate Variability for Real-Time Human-Centered Adaptive Automation Design

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Prinzel, Lawrence J., III; Parasuraman, Raja; Freeman, Frederick G.; Scerbo, Mark W.; Mikulka, Peter J.; and Pope, Alan T.

5d. PROJECT NUMBER

727-05-50-01

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-18316

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)

NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NASA/TP-2003-212442

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 54
Availability: NASA CASI (301) 621-0390 Distribution: Standard

13. SUPPLEMENTARY NOTES

Prinzel and Pope, Langley Research Center, Hampton, VA. Parasuraman, Catholic University, Washington, DC. Freeman, Scerbo, and Mikulka, Old Dominion University, Norfolk, VA.
An electronic version can be found at http://techreports.larc.nasa.gov/ltrs/ or http://ntrs.nasa.gov

14. ABSTRACT

Adaptive automation represents an advanced form of human-centered automation design. The approach to automation provides for real-time and model-based assessments of human-automation interaction, determines whether the human has entered into a "hazardous state of awareness" and then modulates the task environment to keep the operator "in-the-loop", while maintaining an optimal state of task engagement and mental alertness. Because adaptive automation has not matured, numerous challenges remain, including what the criteria are, for determining when adaptive aiding and adaptive function allocation should take place. Human factors experts in the area have suggested a number of measures including the use of psychophysiology. The NASA Technical Paper reports on three experiments that examined the psychophysiological measures of event-related potentials, electroencephalogram, and heart-rate variability for real-time adaptive automation. The results of the experiments confirm the efficacy of these measures for use in both a developmental and operational role for adaptive automation design. The implications of these results and future directions for psychophysiology and human-centered automation design are discussed.

15. SUBJECT TERMS

Automation; Adaptive; Electroencephalogram; Heart-rate variability; Event-related potentials; Human factors

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	70	(301) 621-0390